Communication and learning

Published version:

DOI:10.1093/restud/rdr031

Communication and Learning*

Luca Anderlini
(Georgetown University)

Dino Gerardi
(Collegio Carlo Alberto)

Roger Lagunoff
(Georgetown University)

June 2011

Abstract. We study strategic information transmission in an organization consisting of an infinite sequence of individual decision makers. Each decision maker chooses an action and receives an informative but imperfect signal of the once-and-for-all realization of an unobserved state. The state affects all individuals' preferences over present and future decisions. Decision makers do not directly observe the realized signals or actions of their predecessors. Instead, they must rely on cheap-talk messages in order to accumulate information about the state. Each decision maker is therefore both a receiver of information with respect to his decision, and a sender with respect to all future decisions.

We show that if preferences are not perfectly aligned "full learning" equilibria — ones in which the individuals' posterior beliefs eventually place full weight on the true state — do not exist. This is so both in the case of private communication, in which each individual only hears the message of his immediate predecessor, and in the case of public communication, in which a decision maker hears the message of all his predecessors. Surprisingly, in the latter case full learning may be impossible even in the limit as all members of the organization become perfectly patient. We also consider the case where all individuals have access to a mediator who can work across time periods arbitrarily far apart. In this case full learning equilibria exist.

JEL Classification: C70, C72, C73, D80, D83.
Keywords: Communication, Learning, Dynamic Strategic Information Transmission.
Address for correspondence: Luca Anderlini, Georgetown University, 37^{th} and O Streets NW, Washington DC 20057, USA. `la2@georgetown.edu`

*We are grateful to Bruno Biais, three anonymous referees, Axel Anderson, Leonardo Felli, Johannes Hörner, Navin Kartik, Qingmin Liu, George Mailath, Stephen Morris, Marco Ottaviani, Alessandro Pavan, Larry Samuelson, Leeat Yariv and numerous seminar audiences for helpful comments. Any remaining errors are our own responsibility. All three authors are grateful to the National Science Foundation for financial support (Grant SES-0617789).

1. Introduction

This paper studies the strategic side of information transmission in a large group of agents with imperfectly aligned interests. Our interest is in how and whether intertemporal learning occurs through this transmission.

We posit a model of an ongoing organization consisting of a sequence of imperfectly informed decision makers, each of whom receives information from his predecessors, privately observes some information on his own, then chooses what if any of this information to pass on to future individuals in the sequence. The model we analyze fits well a large firm, a lobbying concern, or a non-profit organization where information evolves through time and is fragmented across a large number of participants. The organization is run by a sequence of decision makers (say CEOs, lead-lobbyists, or managers). Each decision maker in the sequence is forward looking; his payoff is affected by his own and all future decisions. Each of the successive decision makers takes a hidden action, observes some piece of information about an underlying state that affects everyone in the sequence, and then communicates the current state of organizational knowledge to decision makers down the line.

We emphasize that the structure outlined here is not simply a repeated version of the classical "sender-receiver game" of Crawford and Sobel (1982) (henceforth CS). Each of our players is *both* a sender and a receiver, and this changes the framework dramatically.[1]

The privately observed information in the model takes the form of an informative signal of an underlying, decision-relevant state. The underlying state determines which of several activities/tasks (types of competitive strategy for a CEO, degrees of aggressiveness for a lead-lobbyist, types of fund-raising for a non-profit manager) is more rewarding. Each decision maker can then be seen as having a choice of how much time/effort to expend across these multiple activities/tasks. More accurate information about the state therefore improves this choice and benefits everyone in the organization. The critical feature we consider in these types of organizations is that the current decision maker disproportionately bears the different costs of the activities/tasks. In turn, this biases his short run behavior away from the long run common interest of the organization, even if his preferences and those of the entire lineage of decision makers are aligned in the "long run."

Both the state and the players' signals are binary, and the state determines which action is best. Preferences take the form of a quadratic loss around the (unknown) best action given the state. Other things being equal, the current decision maker prefers an action

[1] We return to CS and other related papers in Section 2 below.

that is "lower" than his predecessors would like. Nevertheless, because the signals observed by the various decision makers are (conditionally) i.i.d. across time, if all signals were all publicly observed, then this organization would eventually "fully learn" the underlying truth, resulting in improved decision making for the organization. Our main results show, however, that despite the increasingly accurate information within the organization as a whole, the strategic effects of its dispersion will render full learning impossible.

We examine two canonical "communication technologies," through which the players' messages are viewed by subsequent decision makers. In the first case, communication is *public*. Each player's message is directly observed by all subsequent players in the sequence. Messages, once "published," are difficult or impossible to destroy or manipulate.

In the second case, communication is *private*. Each player's message is directly observed only by the player who immediately follows in the sequence. Messages are easy to hide, manipulate, or falsify. Consequently, each player can act as a "gate-keeper," choosing how much of his knowledge is passed down the line.

In each of these two cases, a natural benchmark to consider in our model is that of truth-telling strategy profiles. In a truth-telling profile, all players reveal truthfully their signal to subsequent players (and pass on truthfully the information they receive from previous players in the case of private communication). Given our i.i.d. signal structure, it is obvious that in this case the organization would fully learn the state. As we move down the sequence, the players' beliefs would necessarily eventually place probability one on the true state.

We show that, both under private and under public communication, not only truth-telling is ruled out, but in fact there are *no equilibria* in which *full learning* occurs. Indeed, in the public communication model we show that there are interior bounds on posterior beliefs about the state beyond which no learning takes place. Only babbling can occur from that point on. Our no-full learning results hold for any discount factor, and any magnitude of the misalignment in preferences. Regardless of whether communication is public or private, all equilibrium outcomes are (ex-ante) bounded away from the Pareto-efficient one induced by the truth-telling strategy profile.

These "no-full learning" results — which hold for any fixed discount factor — naturally beg the question of whether learning can obtain in the limit, along a sequence of equilibria, as players become more and more patient. We refer to this possibility as "limit learning."

As it turns out, the possibility of limit learning depends on which mode of communication, public or private, is considered. We show that limit learning is possible in the private communication case when the preference misalignment is not too large relative to the precision of the signal. By contrast, we show that limit learning is not possible in the public communication model if the signal precision is not too large. The pattern that seems to emerge from these results — parameter configurations in which limit learning is possible with private communication and impossible with public communication — is that "more" communication does not necessarily facilitate learning.

Both the public and private communication cases are forms of decentralized communication. Our final result therefore takes up the cases of *mediated* communication. That is, all players only send messages to a central mediator, who then recommends actions to each on the basis of all messages up that point in the sequence. Under mediated communication, the possibility of full learning is restored. Full learning equilibria do exist in the mediated communication model when the bias is not large relative to the signal's precision, regardless of how heavily the individuals discount the future.

The fact that full learning can obtain in this case, sheds light on the seemingly odd comparison of the private and public communication cases above. In fact, private communication allows for limited mediation in the information flow. Communication between any player and any subsequent one, can effectively be mediated by the intervening players. Our results therefore indicate that "filtered" information being passed down the line facilitates learning.

Going back to our leading examples above, the role of information filtering has implications for the design of available information channels in a firm, a lobbying concern or a non-profit organization. To facilitate learning, the information channels of the organization should allow the CEO, lead-lobbyist or manager to filter the information that is passed down the line, and should not compel him to announce publicly his "report" about the information he has. Rather, in the absence of an unlikely technology that mimics the mediator described above, each agent in charge should be allowed to send a private report directly to his successor, who, in turn, can choose to filter what is sent on.

The rest of the paper is organized as follows. Section 2 reviews some related literature. Section 3 specifies the model and equilibrium concept. There, we define full learning formally. Section 4 gives a very basic intuition for our no-full learning results and outlines some of the difficulties that arise in proving them. Sections 5 and 6 contain our main results, asserting that full learning is impossible both under private and under public communication. Section 7 is concerned with the limit learning results we mentioned above. Section 8 considers the case of mediated communication. Section 9 concludes. The last Section is an Appendix which contains the proofs of the main results.[2]

2. Related Literature

The present paper has elements in common with both the social learning literature and the literature on strategic cheap talk. Both literatures examine why full learning might or might not occur. The possible updating from accumulated informative signals in our model is akin to the social learning literature which tends to focus on costly information acquisition and/or herding effects as the primary sources of incomplete learning. Our costless messages are as in the cheap talk literature which is more concerned with agents' incentives for withholding or manipulating information through the strategic use of messages.

Each of these literatures is too vast to attempt a reasonable summary. Instead, we refer the interested reader to Gale (1996), Sobel (2000), and Moscarini, Ottaviani, and Smith (1998) for surveys and references on social learning, and to Farrell and Rabin (1996), Sobel (2009), Krishna and Morgan (2001, 2008), and Ottaviani and Sørensen (2006) for surveys and references on strategic communication. Here, we limit ourselves to highlighting some similarities and differences with a few contributions that we feel are more closely related to our work.

The social learning side of our model is deliberately kept trivial in order to concentrate on the strategic role of cheap talk. Crucially, past actions are not observed in the present model. More importantly, most of the social learning literature differs from the present paper by assuming that the linkages between agents are purely informational. The standard assumption is that the decisions of one individual do not directly affect the payoffs of another (Banerjee, 1992, among many others). A notable exception is a recent paper by Smith and Sørensen (2008) who examine herding with altruistically linked agents in a setup somewhat similar to ours. Each agent cares about how his action affects the beliefs of future decision makers, thereby adding a signaling component to the model. The result is a tendency for each decision maker to skew one's action against one's belief ("contrarianism"), and this incentive becomes stronger the more certain are the agents about the true state.

The present model shares some features with Smith and Sørensen (2008), notably the presence of dynastic or altruistic linkages between agents, a common value of the unobserved state, and, significantly, the potential to learn from others. The key differences are that, here,

[2]For the sake of brevity, some details of the proofs have been omitted from the Appendix as well. These can be found in a supplement to the paper available at http://www.anderlini.net/learning-omitted-proofs.pdf.

a payoff bias is assumed between different generations of decision makers, and the focus is on learning from costless messages rather than costly actions. Smith and Sørensen (2008) study an equivalent team problem (all individuals maximize the same average discounted sum) in which the externalities that generate herding are internalized. They show that under some conditions this is equivalent to a planner's experimentation problem with incomplete information.

Payoff bias and cheap talk go hand in hand in the present set-up. While these two features are not necessarily inseparable — Kartik (2009) incorporates lying costs into the CS model thereby bridging the gap between costly signaling and cheap talk — our motivation is in isolating incentive effects of intergenerational learning via pure cheap talk.[3]

Within the cheap talk literature, our model is closer to those with a dynamic or dynastic element. Spector (2000), for instance, examines a repeated game of communication in a large population with a multi-dimensional policy space. Everyone in the population shares the same underlying policy preferences. The population is partitioned into two groups characterized by different priors about an unobserved state. Each individual receives an informative signal of the state. The two groups alternate in the role of senders and receivers with a new sender and receiver randomly drawn from the appropriate group each period.

Unlike ours, Spector's players are myopic. Hence he does not look at the dynamic component of strategic information transmission. Rather, his focus is on the effect of conflict on a multi-dimensional policy space which, surprisingly, in a steady-state reduces to at most a single-dimensional disagreement.

Learning in various dynamic versions of the basic sender-receiver game is studied in Taub (1997), Carrillo and Mariotti (2000), Li (2007), and Golosov, Skreta, Tsyvinski, and Wilson (2011). In Taub (1997), the ideal policy positions of the participants vary across time according to serially correlated processes that are uncorrelated with one another. Both Taub's model and ours involve learning via accumulated messages. In his model, however, partial learning is possible due to the serial correlation in the sender's bias. In ours whatever learning occurs takes place automatically, provided information is properly transmitted through time.

Carrillo and Mariotti (2000) examine the learning decision of a single agent with present-biased preferences. In each period, the agent chooses the precision of a public signal. Given

[3]Kartik's model shows, among other things, that lying costs turn cheap talk into costly signaling. He shows that when the sender's bias makes him prefer higher actions relative to the receiver, then full separation is possible in the lower regions of the type space. This is not possible in the standard CS model with a continuum of types.

the bias, the problem may be re-interpreted as a dynastic game similar to our public communication setup. They show that both complete and incomplete learning equilibria exist.

Li (2007) considers the sequential arrival of messages which may reveal information about the sender's *ability* (the quality of his signals). Intriguingly, she finds that a sender who "changes his mind" may be a signal of high ability. In our set up all signals have the same precision since they are i.i.d. and hence there is nothing to infer about the sender's ability in the sense of Li (2007).

Golosov, Skreta, Tsyvinski, and Wilson (2011) set up their model close to the original CS framework. A state is realized once and for all and observed by the sender. Then, in each of a finite number of periods, the sender sends a message to the same receiver who takes an action each time. They show that full information revelation is possible.

Finally, this paper builds on our previous work on dynastic communication in Anderlini, Gerardi, and Lagunoff (2008, 2010), Anderlini and Lagunoff (2005), and Lagunoff (2006).[4] These papers consider dynastic games in which messages are about the past history of play. Because there are no "objective" types (i.e., no exogenous states or payoff types), there is no learning in the traditional sense of the word. These contributions focus on whether social memory, embodied in the beliefs of the new entrants, is accurate, and what pitfalls may befall a society if it is not.

3. The Framework

3.1. Model

Time is discrete, and indexed by $t = 0, 1, 2, \ldots$ In period 0 Nature selects the state ω which can be either 1 or 0. State $\omega = 1$ is chosen with probability $r \in (0, 1)$. Nothing else takes place in period 0, and no-one in the model observes Nature's choice of ω, which is determined once and for all.

There is a period-by-period flow of imperfect information about the state ω. We take this to have the simplest form that will allow for an interesting model: a sequence of imperfect but informative conditionally i.i.d. signals. The realized signal in period t is denoted s_t, and its symmetric structure is parameterized by a single number $p \in (1/2, 1)$ as

$$\Pr\left\{ s_t = 1 \;\middle|\; \omega = 1 \right\} = \Pr\left\{ s_t = 0 \;\middle|\; \omega = 0 \right\} = p \tag{1}$$

[4] Related models are also found in Kobayashi (2007) and Lagunoff and Matsui (2004).

With obvious terminology, we refer to p as the signal's quality.[5]

There is a countable infinity of decision makers, each indexed by $t = 1, 2, \ldots$ Each decision maker's tenure lasts for one period; the decision maker in period t chooses an action $a_t \in \mathbb{R}$ at time t, but he also derives state-dependent utility not only from his own choice but also from all the choices $a_{t+\tau}$, with $\tau = 1, 2, \ldots$, of all his successors. We refer to this decision maker simply as "player t."

We work with a specific payoff function that captures our desiderata and at the same time ensures tractability.[6] Fix a sequence of actions $a_1, \ldots, a_t, \ldots$ and a state ω. The payoff of player t is written as

$$-(1-\delta)\left[\sum_{\tau=0}^{\infty} \delta^{\tau}(\omega - a_{t+\tau})^2 + 2\beta a_t\right] \tag{2}$$

with δ a common discount factor in $(0, 1)$. The squared term in (2) is a quadratic loss function that reaches a maximum at $a_t = \omega$ for $\omega = 0, 1$. The term $2\beta a_t$ in (2) embodies the bias in the decision makers' preferences. Each player t bears an extra "effort" cost $2\beta a_t$, increasing in a_t.[7] Aside from this, for given beliefs, the decision makers' preferences are aligned. Of course, what matters to player t is the expected value of (2), given his beliefs about ω and the sequence of all future actions $\{a_{t+\tau}\}_{\tau=1}^{\infty}$.

Because of the way the bias term is parameterized by β, it is immediate to check that if player t believes that $\omega = 1$ with probability x_t, then his choice of a_t to maximize (2) is simply

$$a_t = x_t - \beta \tag{3}$$

Given (3), we refer to β simply as *the* bias. Throughout, we assume that $\beta \in (0, 1/2)$.[8]

Player t's information comes from his predecessor(s) in the form of payoff irrelevant messages. He does not observe directly any part of the past history of play or realized signals.

[5]While the assumption of conditionally i.i.d. signals is important for our arguments, the symmetry of the signal structure could easily be dispensed with.

[6]Subsection 9.1 discusses possible generalizations of payoffs.

[7]We can also interpret each a_t (appropriately normalizing so that $a_t \in [0, 1]$) as the fraction of time that t spends on one of two tasks available, with each task yielding different losses in the two states. The bias term is then interpreted as embodying the fact that the task on which the fraction of time a_t is spent is more costly for t than the others.

[8]The restriction $\beta < 1/2$ simplifies some of our computations, and only disposes of uninteresting cases.

In the case of *private communication* player t observes a message m_{t-1} sent by player $t-1$, which is observed by no-one else. In the case of *public communication* player t observes all messages $m_{t-\ell}$ (with $\ell = 1, \ldots, t-1$) sent by all preceding players. It is useful to establish a piece of notation for what player t observes by way of messages that encompasses both the private and public communication cases. We let $m^{t-1} = m_{t-1}$ in the private communication case, while $m^{t-1} = (m_1, \ldots, m_{t-1})$ in the public communication case. (Set $m_0 = m^0 = \emptyset$.)

We establish our results for any choice of finite message space.[9] Here, we simply establish some basic notation. The message space available to each player t from which to pick m_t is denoted by $M_t \neq \emptyset$, with $M_0 = \emptyset$. We also let M represent the entire collection $\{M_t\}_{t=0}^{\infty}$ of message spaces available in each period.

After observing m^{t-1}, player t picks an action a_t. After choosing a_t he observes the informative but imperfect signal s_t, and after observing s_t he decides which message m_t to send.[10] Note that, crucially, player t does not learn the state before sending m_t. This can be interpreted as his period t payoff simply accruing after he sends m_t, or as his experiencing a payoff that depends on his action a_t and the signal s_t in an appropriate way.[11]

Given this time line, player t's strategy has two components. One that returns a probability distribution over actions a_t as a function of m^{t-1}, and one that returns a probability distribution over messages m_t as a function of m^{t-1} and s_t. We denote the first by λ_t and the second by σ_t.

Notice that there is a sense in which the λ_t part of t's strategy is trivial. Given his beginning-of-period belief $x_t \in [0,1]$ that $\omega = 1$, using (2) and (3) we know that in any equilibrium it will be the case that $a_t = x_t - \beta$ with probability one. Recall also that a_t is not observed by any player other than t himself.

Since there is no loss of generality in considering a λ_t that returns a pure action, we save on notation by *defining* it as returning just a real number $a_t \in \mathbb{R}$. In the case of messages however, mixing must be explicitly considered. Hence we let $\Delta(M_t)$ be the set of all possible probability distributions over M_t, and let this be the range of σ_t. Formally

$$\lambda_t : M^{t-1} \to \mathbb{R} \ \text{ and } \ \sigma_t : M^{t-1} \times \{0,1\} \to \Delta(M_t) \tag{4}$$

[9] The results do not depend on the finiteness of the message space. We make the assumption purely to ease the technical burden.

[10] The fact that s_t is observed after the choice of a_t simplifies the analysis but is inessential to our results.

[11] The latter would be the analogue of what is routinely done in the large literature on repeated games with imperfect private monitoring. See Mailath and Samuelson (2006) for an overview and many references.

so that $\sigma_t(m^{t-1}, s_t)$ is player t's mixed strategy in messages given t's observed message history m^{t-1} and his signal s_t. Occasionally we write $m_t = \sigma_t(m^{t-1}, s_t)$ to mean that t sends message m_t with probability one.

Since the only way to transmit information are the messages described above, it is clear that x_t is entirely determined by the message-sending behavior of t's predecessors $(\sigma_1, \ldots, \sigma_{t-1})$ together with the observed m^{t-1}. Thus, all strategic behavior (and hence any equilibrium) is entirely pinned down by a profile $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_t, \ldots)$.

To avoid unnecessary use of notation, and since this does not cause ambiguity, from now on we will refer to $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_t, \ldots)$ as a strategy profile, leaving it as understood that the associated $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_t, \ldots)$, determining actions via (2) and (3), is to be considered as well. Of course, actions and messages sent in equilibrium will be determined by (2), (3) and the players' on- and off-path beliefs. Luckily, off-path beliefs need not be specified with any piece of notation. The reason is simply that off-path beliefs will not play any role in our results. Since the on-path beliefs are entirely pinned down by $\boldsymbol{\sigma}$ via Bayes' rule, referring to this profile will always suffice.

Throughout the paper, an "Equilibrium" will refer to a Weak Perfect-Bayesian Equilibrium (henceforth WPBE) of the game at hand (see for instance Fudenberg and Tirole, 1991). This is the weakest equilibrium concept that embodies sequential rationality that applies to our set up. Our main focus is on ruling out certain outcomes as possible in any equilibrium. Therefore, a weaker equilibrium concept strengthens our main results. It is in fact also the case that whenever we construct an equilibrium in any of our arguments below (Propositions 3 and 5 below), then the equilibrium will also satisfy the more stringent requirements of a Sequential Equilibrium in the sense of Kreps and Wilson (1982).

3.2. Learning

Our main question is whether the accumulated information in this organization will suffice to eventually uncover, or fully learn, the true state. Hence, we now need to be precise about what we mean by full learning in our model.

Fix a strategy profile $\boldsymbol{\sigma}$. As we noted above, the belief of player t at the beginning of period t that $\omega = 1$ is entirely determined by $\boldsymbol{\sigma}$ and the m^{t-1} he observes. Denote it by $x_t(\boldsymbol{\sigma}, m^{t-1})$ with the dependence on $\boldsymbol{\sigma}$ or m^{t-1}, or both, sometimes suppressed as convenient.

Now fix a realization $\mathbf{s}$ of the entire stochastic process $\{s_t\}_{t=1}^{\infty}$. Via the strategy profile $\boldsymbol{\sigma}$, the realization $\mathbf{s}$ clearly determines m^{t-1} for every $t = 1, 2, \ldots$ This makes it clear that a given profile $\boldsymbol{\sigma}$ (together with a realization of the state ω and the signal structure (1) parameterized by p) defines a stochastic process $\{x_t\}_{t=1}^{\infty}$ governing the (on-path) evolution of the players' posterior. The probabilities (conditional when needed) of events in this process will be indicated by $\Pr_{\boldsymbol{\sigma}}$.[12]

Definition 1. *Full Learning: Fix p, β, r, δ, and a collection of message spaces M. We say that a profile $\boldsymbol{\sigma}$ induces full learning of the state if and only if for every $\varepsilon > 0$ there exists a $\bar{t} \geq 1$ such that $t > \bar{t}$ implies*

$$\Pr_{\boldsymbol{\sigma}}\left\{x_t > 1-\varepsilon \;\middle|\; \omega = 1\right\} > 1-\varepsilon \qquad \text{and} \qquad \Pr_{\boldsymbol{\sigma}}\left\{x_t < \varepsilon \;\middle|\; \omega = 0\right\} > 1-\varepsilon \tag{5}$$

In words, $\boldsymbol{\sigma}$ induces full learning of the state if the associated stochastic process $\{x_t\}_{t=1}^{\infty}$ converges in probability to the true state ω.

Our first two results, Propositions 1 and 2 below, establish that in equilibrium full learning is impossible both in the private and public communication environments.

However, Propositions 1 and 2 refer to a *given* equilibrium for a given set of parameters. Since in the limit as δ approaches 1 the players' preferences become aligned, the natural next question is whether learning can obtain in the limit as δ approaches 1, along a *sequence of equilibria*.

To examine this claim, we define formally the notion of "limit learning" in our model.

Definition 2. *Limit Learning: Fix p, β and r. We say that limit learning of the state is possible if and only if for every $\varepsilon > 0$ we can find a $\bar{t} \geq 1$ and a $\overline{\delta} < 1$ such that for every $\delta > \overline{\delta}$ there is a collection of message spaces M and an equilibrium $\boldsymbol{\sigma}^*(\delta)$ that satisfies*

$$\Pr_{\boldsymbol{\sigma}^*(\delta)}\left\{x_t > 1-\varepsilon \;\middle|\; \omega = 1\right\} > 1-\varepsilon \qquad \text{and} \qquad \Pr_{\boldsymbol{\sigma}^*(\delta)}\left\{x_t < \varepsilon \;\middle|\; \omega = 0\right\} > 1-\varepsilon$$

for every $t > \bar{t}$.

4. Some Bare-Bones Intuition

This section develops some of the basic intuition underlying our main results, particularly the impossibility results of Propositions 1 and 2. We highlight some of the difficulties that the

[12] It is clear that $\boldsymbol{\sigma}$ also induces a stochastic process $\{m_t\}_{t=0}^{\infty}$ governing the (on-path) evolution of messages sent by each player. With a slight abuse of notation the probabilities (conditional when needed) of events in this process will be indicated in the same way by $\Pr_{\boldsymbol{\sigma}}(\cdot)$. The argument of the operator will ensure that this does not cause any ambiguity. Finally, because $\boldsymbol{\sigma}$ pins down the on-path evolution of both $\{x_t\}_{t=1}^{\infty}$ and $\{m_t\}_{t=0}^{\infty}$, it also induces *updated* probabilities that ω is one or zero, conditional on any (positive probability) event concerning $\{x_t\}_{t=1}^{\infty}$ or $\{m_t\}_{t=0}^{\infty}$, or both. These will also be identified using the notation $\Pr_{\boldsymbol{\sigma}}(\cdot)$, with the arguments used to avoid any ambiguity.

arguments run into, and outline how these are resolved in very broad terms. More detailed sketches of the actual arguments follow each proposition below.

4.1. Looking One Period Ahead

To see that the bias in preferences may make the truthful revelation of information problematic just from one period to the next, one can reason in much the same way as in the CS world. Consider just a single individual (sender) and his successor (receiver). Let x be the initial belief (that the state is equal to one) of the sender. Suppose first that the sender's strategy is to reveal his signal $s \in \{0, 1\}$. Since the signal is informative, the receiver will now have beliefs either x^0 (if the report is $s = 0$) or x^1 (if the report is $s = 1$) that satisfy $x^0 < x < x^1$. Because of the bias, the receiver will take action $x^0 - \beta$ in one case and $x^1 - \beta$ in the other. Suppose on the other hand that the sender babbles and reveals no information. For instance, he reports $s = 1$ regardless of what he really observes. Then the receiver's belief is the same as the sender's before he observes s, namely x. Hence the receiver takes action $x - \beta$.

For the information to be truthfully transmitted between sender and receiver we need to satisfy the usual two incentive compatibility constraints. The sender must *both* (weakly) prefer action $x^1 - \beta$ to action $x^0 - \beta$ when his belief is x^1 *and* he must (weakly) prefer action $x^0 - \beta$ to action $x^1 - \beta$ when his belief is x^0. The first constraint does not bind since actions of the receiver are always skewed downward.

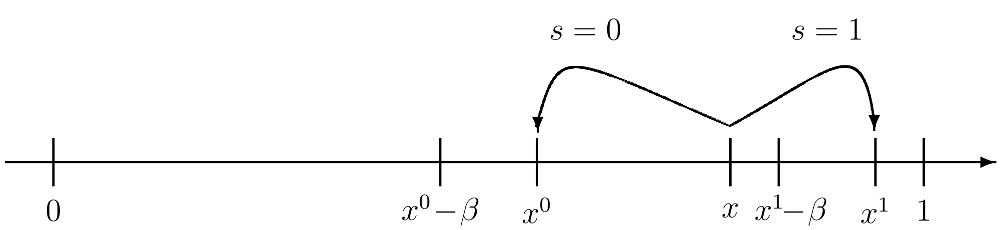

Figure 1: One Period Incentive Compatibility

The second constraint on the other hand might be impossible to satisfy. This constraint is satisfied if $x^0 - \beta$ is no further from x^0 than is $x^1 - \beta$, as is for instance the case in Figure 1. Since the constraint can be written as $2\beta < x^1 - x^0$, it is clear that if $x^1 - x^0$ is small relative to β then the second incentive compatibility constraint cannot be satisfied. If this is the case, after observing $s = 0$, the sender prefers the action that the receiver would take if he believed the sender's message and was told that $s = 1$. Hence there is no equilibrium in which the sender's messages fully reveal the signal s.

It is critical to notice that the possibility of truthful revelation just one period ahead depends on the *relative* size of $x^1 - x^0$ and β. Given this, it is immediate that one period ahead truth-telling will be possible when the belief x takes "intermediate" values (provided that β is not too large), while it will become impossible as x approaches either zero or one. This is simply because Bayes' rule tells us that the updating effect will be larger for intermediate values of x and will shrink all the way to zero as x goes to the extremes of $[0, 1]$.

4.2. Difficulties Of The Full-Blown Dynamic Case

Armed with the observations of Subsection 4.1, now consider the full-blown dynamic model we described in Section 3. To fix ideas, suppose communication is public (similar issues arise when communication is private). Suppose that $\omega = 1$ is the true state and that we had an equilibrium with perpetual truth-telling. Of course, truth-telling is particularly simple way to achieve full learning. Then, in equilibrium, the beginning-of-period belief of player t, x_t, and the updated beliefs, x_t^0 and x_t^1, following signals 0 and 1 (resp.), all must converge to one with probability one. This of course means that the difference $x_t^1 - x_t^0$ must eventually become arbitrarily small. So eventually in this putative truthful equilibrium we must be in a situation in which $2\beta > x_t^1 - x_t^0$. From the discussion of the one-period case, we know that a player t who finds that $2\beta > x_t^1 - x_t^0$ and observes a signal $s_t = 0$ will have higher payoff in period $t+1$ if he deviates and reports $s_t = 1$. However, before deciding whether to deviate from the putative truth-telling equilibrium, player t will have to examine his incentives beyond $t + 1$.

Consequently, consider t's payoff in any future period $t+\tau$, with τ possibly large. Taking the truthful strategies of the others as given, a deviation to $s_t = 1$ when in fact $s_t = 0$ will simply take the entire sequence of beliefs from $t + 1$ to $t + \tau$ up two "notches" relative to where it would be if t did not deviate (in the same sense that x_t^0 is two notches down from x_t^1 — it takes two realizations of the high signal for Bayes' rule to take beliefs from x_t^0 to x_t^1).

Consider a sequence of signal realizations that takes the beliefs gradually towards the center of the interval $[0, 1]$. As we noted before, the one-period-ahead calculation for a player with intermediate beliefs tells that player *not to deviate* from truthful revelation, provided that β is not too large. So, with some probability, the deviation at t has the same effect in period $t + \tau$ as a deviation by player $t + \tau$ whose beliefs start from an intermediate position. In this case, there would be a loss in $t + \tau$ from the deviation.

Of course, calculating the entire path following a deviation will depend crucially on the signal precision (how fast the beliefs move through time), and on how much the individuals discount the future. The computation is not trivial since both gains and losses from deviating shrinks to zero as x_t goes to either one or zero. Moreover, in order to rule out *any* equilibrium with full learning, we need to look beyond simple truth-telling strategies. For one thing, equilibria in mixed message strategies may exist and must therefore be considered. For another, there could exist equilibria that are not Markov in the sense that they could have continuation paths that differ following different histories leading to the same belief.

In this latter case, one could imagine a scenario in which full learning arises because the equilibrium continuations effectively "punish" deviations from truth-telling behavior. Consider once again a player t who finds that $2\beta > x_t^1 - x_t^0$, and observes a signal $s_t = 0$. Just as before, in a full learning equilibrium this must happen at some point with probability one. Looking one period ahead, player t prefers to mimic the type who observed $s_t = 1$. However, because the equilibrium need not be Markov, the continuation equilibrium could for instance be such that the beliefs get "trapped" in the middle of $[0, 1]$ for long periods (because of prolonged spells of babbling) *only* following the message sent by type $s_t = 1$. So, by deviating, player t of type $s_t = 0$ may be "punished" because the future path of beliefs following the deviation may yield the losses associated with beliefs in the intermediate range of $[0, 1]$ with much higher probability than in the case of a truth-telling strategy profile. Our impossibility results overcome the difficulties we have described mostly by bounding continuation payoffs or their changes in case of deviations in a variety of ways.

4.3. Broad Outline of the Two Proofs: Similarities and Differences

The proofs of Propositions 1 and 2 below share some structure that arises naturally from our intuitive discussion so far. A more detailed outline of each proof will follow the formal statements of results, but in the meantime we describe some key similarities and differences between the two cases.

In both private and public communication, the payoff to player t other than the current-period component is divided into the payoff that accrues to t in period $t + 1$, and the rest of t's continuation payoff, stemming from the actions taken in $t + 2$ and beyond.[13] For the purposes of this discussion we will refer to the former as the payoff one period ahead, and to the latter as the continuation payoff.

In both the private and public communication cases, the argument begins in the following way. If we had any equilibrium with full learning, then for some large t the belief of player t

[13]Dealing with the part of t's payoff that accrues at t — determined by the $\lambda_t(\cdot)$ part of his strategy as in (4) — is trivial as we noticed in Subsection 3.1. It will simply be given by (3) and not affect any of the play that follows.

must be eventually arbitrarily close to 0 (if $\omega = 0$) or to 1 (if $\omega = 1$), *and* player t's message strategy must be sufficiently informative for his successors' beliefs to continue on a convergent path to the true state. We refer to this informative strategy as "truth-telling" (although the term is somewhat inaccurate since, for instance, the strategy could be mixed).

At this point, recall the observations we made in Subsection 4.1 above. There, we showed that for a given β, player t's one period ahead payoff is larger if he deviates from truth-telling. In order to rule out all equilibria with full learning, however, it must also be the case that the one period gain from deviating outweighs the possible future loss in the player's continuation payoff. This is the point where the two proofs necessarily differ.

With private communication, any player can "wipe out" the past history of play entirely if he so decides. It might therefore seem easy to negate full learning because a player has so many options when contemplating a deviation — he can in principle send a message to place his successor's beliefs anywhere in $[0, 1]$. However, as our previous discussion indicates, if player t were to deviate from truth-telling in order to realize a one period gain, player $t+1$ can conceivably "punish" t's deviation by sending a message that moves beliefs of future players, and consequently player t's continuation payoff, far away from what they would have been without the deviation. In other word, the players who follow t in the sequence have the same latitude, and so it may be feasible for them to deter a deviation by player t.[14]

Ultimately, one needs to show that these types of "large deterrents" by $t + 1$ — those messages that moves beliefs of future players far from what they would have been without the deviation — are not profitable for $t+1$. The proof establishes this by showing that these types of deterrents create streams of actions that are *also* far from player $t+1$'s own ideal. As a result, any initial deviation by t will induce at most a small departure from t's continuation payoff in equilibrium. The proof shows, in fact, that the difference in continuation payoffs to player t shrink to zero faster than one period ahead gain from lying as his beliefs get arbitrarily close to the true state. This is sufficient to destroy the putative full learning equilibrium.

By contrast, under public communication the "size" of any potential deviation is limited since player t cannot undo the message history — he can only lie about his signal. But this means that the size of player $t + 1$'s deterrent is also limited. Consequently, player $t + 1$ cannot as easily "punish" player t for a deviation. The more limited deterrent should

[14]Clearly, since player $t + 1$ is not even aware of t's deviation he is not literally "punishing" t. Instead, it is the equilibrium construction itself that might prevent deviations by encoding the punishment into $t + 1$'s prescribed message in the continuation.

therefore strengthen the impossibility result: t's message here may have unavoidable effects on the continuation of play and this, in turn, limits the downside from a contemplated deviation. The argument for Proposition 2 therefore proceeds by directly comparing the overall payoff to t when he does not deviate, with the worst possible case over all possible strategy profiles (of other players) if he does deviate. This payoffs difference provides an obvious lower bound to the gains from a deviation under full learning. We therefore set this problem up as a constrained optimization problem and show that its saddle value gives positive gains for the deviation as beliefs get close to the truth.

To summarize, with private communication players can wipe away history, and so deviations appear to be unlimited. However, their impact can be blunted or reversed (in payoffs) by future players' ability to wipe away history as well. By contrast, with public communication, deviations are more limited in size but their effect is more permanent. In either case, it turns out that showing that full learning is impossible is a nontrivial task.

5. No Full Learning with Private Communication

With private communication full learning of the state cannot occur in equilibrium. This is so regardless of the message spaces, discount factor, signal precision, bias or prior.

Proposition 1. *No Full Learning with Private Communication: Fix any M, δ, p, β and r. Then there is no equilibrium $\boldsymbol{\sigma}^*$ of the model with private communication that induces full learning of the state.*

The formal proof is in the Appendix. Here, we outline an informal sketch of the argument. To fix ideas, consider the case where $\omega = 1$ is the true state and suppose, by way of contradiction, that an equilibrium with full learning did exist. Then we know there is some player t whose beginning-of-period belief is arbitrarily close to one.

For illustrative purposes, suppose that player t has messages that can fully reveal his end-of-period belief to player $t+1$. In other words, he has messages m' or m that can fully reveal the end-of-period belief, x' or x, that he would have after observing $s_t = 0$ or $s_t = 1$, respectively.[15] Hence, $x > x'$ and, because we take player t's beginning-of-period belief to be arbitrarily close to one, the messages m and m' will cause x and x' to be so as well.

[15] In general, of course, no such messages need exist in the private communication environment. While player t may indeed send different messages after observing $s_t = 0$ and $s_t = 1$, resp., these messages may fail to transmit his end-of-period belief to player $t+1$. This is because other "types" of player t may be using the same messages on the equilibrium path. Of course, in the Appendix we treat an exhaustive set of possible cases.

Suppose that player t observes $s_t = 0$. He is of course supposed to send message m' to player $t+1$, thus inducing $t+1$ to have belief x'. What are the incentives of player t to deviate by sending m instead? It turns out that a lower bound to the net gain from this deviation is given by

$$(x - x')\,\beta \; - \delta(x - x')\,A$$

The first part $(x-x')\,\beta$ bounds from below the net gain in the period $t+1$ payoff to player t. Looking at this period $t+1$ payoff in isolation, player t would be better off by deviating and sending message m. This is for exactly the same reason that we pointed out in Subsection 4.1 above, with the help of Figure 1, and follows directly from the fact that $x - x'$ is small.

The second part consists of the potential loss from the deviation in the stream of payoffs that accrue from $t+2$ onward. This stream coincides with the continuation payoffs of player t's successor, player $t+1$. We show that the difference (positive or negative) is bounded below by $-(x - x')\,A$, with A a term that depends on the expected continuation payoffs computed after player $t+1$'s action has been chosen. The critical step is in showing that as $x - x'$ shrinks to zero, the term A in the bound *also shrinks to zero*. This is not hard to show when the continuation payoff is conditioned on $\omega = 1$. Using the incentive constraints of player $t+1$, it must be that the continuation payoffs are uniformly close to each other since the conditioning event ($\omega = 1$) has probability close to one in that player's beliefs. If the payoffs were not close to each other one of the two "types" (either the one receiving m or the one receiving m') would necessarily have an incentive to behave like the other.

To show that the continuation payoffs are close to each other when conditioning on $\omega = 0$ involves an extra difficulty. The event $\omega = 0$ has probability close to zero in player $t+1$'s beliefs, hence the differences in conditional payoffs have a negligible effect on his incentives. We circumvent this difficulty by showing that the incentives of subsequent players from $t+2$ onward will not lead to divergent equilibrium continuation paths in expectation.[16]

Since β is fixed and A converges to zero as both x and x' converge to one (and to each other), the net gain from deviation after signal $s_t = 0$ is observed is positive. This destroys the putative equilibrium in which full learning occurs.[17]

[16]Briefly, we show that if the payoffs associated with two continuation paths conditional on $\omega = 0$ are far apart, then the payoffs associated with those realized paths conditional on $\omega = 1$ *must also be far apart.* This in turn contradicts the uniform closeness of continuation payoffs when $\omega = 1$ as established above.

[17] With relatively minor modifications to the proof, one can show that there does not exist a sequence of equilibria that yield outcomes that are closer and closer to full learning. Formally, fix any δ, p, β and r. Then the set of ε for which any equilibrium satisfies (5) is bounded away from zero.

6. No Full Learning with Public Communication

Just as in the case of private communication, with public communication full learning of the state cannot occur in equilibrium. In this case too, this is so regardless of the message spaces, discount factor, signal precision, bias or prior.

Proposition 2. *No Full Learning with Public Communication: Fix any M, δ, p, β and r. Then there is no equilibrium $\boldsymbol{\sigma}^*$ of the model with public communication that induces full learning of the state.*

The conclusion of Proposition 2 is the same as for Proposition 1: posterior beliefs do not converge in probability to the true state. In this case, this is true despite the fact that decision makers can access the entire history of messages that precedes them. Here we give an informal sketch of the argument and place the full proof on the Appendix.

As with Proposition 1, the proof proceeds by contradiction: suppose there is an equilibrium $\boldsymbol{\sigma}^*$ with full learning. We first show that for any interval of beginning-of-period beliefs close to one, there is at least one such belief at which the supports of the mixed message strategies cannot coincide. This step follows, roughly speaking, from indifference conditions under mixed strategies.

In turn, this means that one of two cases must hold for a date t sender with beliefs close to one. Either (a) he has a message m_t that makes recipient $t+1$ certain that $s_t = 1$ (since the message would lie outside the support of $\sigma_t^*(m^{t-1}, s_t = 0)$), or (b) he has a message m_t' he can send that makes the recipient certain that $s_t = 0$.

Let $V(x, y, \boldsymbol{\sigma}^*)$ denote the continuation payoff, evaluated at an arbitrary message strategy profile $\boldsymbol{\sigma}$, to any sender whose belief at the message stage is x and who sends a message that induces belief y by the recipient. To fix ideas, consider case (a) above: there is a message history m^{t-1} and a message m_t that makes the recipient $t+1$ certain that $s_t = 1$. Consider the incentives of the sender if, in fact, he actually observes $s_t = 0$. The message m_t would then be a deviation. Let x_t denote the sender's beliefs after he observes signal $s_t = 0$. Note that $x_t < z(x_t)$ where $z(x_t)$ is t's beginning-of-period belief before observing any signal. Clearly, the signal $s_t = 0$ would lower t's belief that $\omega = 1$. Now, because we have supposed an equilibrium with full learning, t's messages would be at least weakly informative in the sense that $x_t \leq y_{t+1} < z(x_t)$ where y_{t+1} is the belief of player $t+1$ after a message in the equilibrium support of $\sigma_t^*(m^{t-1}, s_t = 0)$. If this sender were to deviate and send m_t instead then the recipient's belief would be $h(x_t) \equiv p^2 x_t / [p^2 x_t + (1 - x_t)(1 - p^2)] > z(x_t)$. In this case, the gain from a deviation to m_t is clearly bounded below by

$$\min_{y,\boldsymbol{\sigma},\boldsymbol{\sigma'}} \quad V(x_t, h(x_t), \boldsymbol{\sigma}) - V(x_t, y, \boldsymbol{\sigma'}) \quad \text{subject to} \quad x_t \leq y \leq z(x_t)$$

The main step in the proof characterizes the solution to this constrained optimization program. Specifically, we apply an Envelope Theorem of Milgrom and Segal (2002) in order to compute the saddle value in a neighborhood of $x_t = 1$. We then show that its value is strictly positive, indicating that m_t is a profitable deviation. Consequently, a full learning equilibrium cannot exist.

Finally, we note the lower bound on the gain from deviation is independent across all equilibria. This is clear from the constrained problem above. Therefore in any equilibrium there are bounds on beliefs beyond which no learning takes place. Only babbling can occur if they are reached. Of course, this also implies that, as in the case of private communication, the set of equilibria is "bounded away" from full learning in this case too.

7. Limit Learning

We now return to the issue of limit learning and investigate whether it is possible in our model. The answer depends on, among other things, the type of communication allowed. In the model with public communication we are able to show that when p is not too large, even limit learning is impossible. On the other hand, our results for private communication suggest that limit learning is possible under some conditions in this case.

7.1. Private Communication

Provided the bias is not too large relative to the signal's quality, limit learning is possible when communication is private.

Proposition 3. *Limit Learning with Private Communication: Fix any r and suppose that $\beta < p - 1/2$. Then limit learning is possible in the model with private communication.*[18]

At this point it is worth mentioning again that, as we alluded to in Subsection 3.1, the limit learning of the statement of Proposition 3 can be sustained with Sequential Equilibria rather than just with WPBE.

[18] In the Appendix we prove a slightly stronger statement. The one given here is phrased to allow direct comparability with the statement of Proposition 4 below. The proof of Proposition 3 in the Appendix yields a collection M of (finite, of course) message spaces and an equilibrium $\boldsymbol{\sigma}^*(\delta)$ satisfying the requirements of Definition 2 which do not in fact depend on δ but only on the value of ε (or equivalently $\overline{\delta}$) in Definition 2.

In the Appendix we construct an appropriate sequence of Sequential Equilibria. As with our previous two results, we only give a sketch here. For the sake of clarity, this is divided into four parts. Throughout this sketch, for simplicity we imagine that the prior r is equal to one half. If this were not the case the review phase we sketch below would not be symmetric making the outline a lot more cumbersome to follow. The role of the condition $\beta < p - 1/2$ would also be lengthier to explain.

The argument begins with two observations. The condition $\beta < p - 1/2$ in the statement of the proposition has an immediate interpretation. This condition tells us that any player with a belief of $1 - p$ or less will strictly prefer action $-\beta$ to action $1 - \beta$ to be taken in the future. Moreover, these preferences will be reversed for any player with a belief of p or more. Since $r = 1/2$ it is clear that the belief of any player who knows (or believes) the history of signals to contain more zeroes than ones will have a belief of $1 - p$ or less, while any player who knows (or believes) the history of signals to contain more ones than zeroes will have a belief of p or more. So, the majority of signals being zero or one will determine a player's strict preferences between future actions $-\beta$ and $1 - \beta$ as above.

The second part of the argument outlines the mechanics of the review phase that characterizes the equilibria we construct to support limit learning. In the first T periods of play, with T a large odd number, equilibrium consists of a review phase. The purpose of the review phase is to establish whether the majority of signals in the first T periods is zero or one. The review phase is said to have outcome zero or one according to which one of these events occur. Of course if the proportion of zeroes or of ones is very large in the initial part of the review phase, its outcome may be determined early.

During the first T periods, unless the outcome has already been determined, the players collectively count the number of signals equal to one that have been realized so far. Each player adds one to his predecessor's message if he observes a signal of one, and leaves the count unchanged if he observes a signal of zero. He then sends the new tally as a message to the next player. If at any point the outcome of the review phase is determined, the current player sends a message that indicates that the outcome of the review phase has been determined and whether it is zero or one. From that point on, up to T and beyond, only that message is sent forward through time each period and thus reaches all subsequent players. No further information is ever added.

The third part of the argument consists of noticing that, using the Weak Law of Large Numbers, making the review phase longer and longer, we can ensure that the beliefs of all players from $T + 1$ onwards are arbitrarily close to the true state (within ε of it), with arbitrarily large probability (at least $1-\varepsilon$) as required for limit learning to occur. Intuitively, this is clear from the fact that the signal is informative and, in equilibrium, all players from $T+1$ onwards only know whether the majority of a large number T of realizations of the signal is zero or one.

The fourth part of the argument consists of the verification that no player wants to deviate from the prescribed behavior. It is useful to consider two cases. Begin with any player $t \geq T+1$, after the review phase is necessarily over. As we noted in the third part of the argument, this player's belief is either arbitrarily close to zero or arbitrarily close to one, depending on the message he receives about the outcome of the review phase. If he deviates and reports an outcome of the review phase different from the one he received, all subsequent actions will be arbitrarily close to $-\beta$ and his belief arbitrarily close to one in one case, and will be arbitrarily close to $1-\beta$ with his belief arbitrarily close to zero in the other case. Therefore, because of what we know from the first part of the argument, no player $t \geq T+1$ stands to gain from deviating from the proposed equilibrium.

Next, consider any player $t \leq T$ when the outcome of the review phase may or may not have been determined. Consider his continuation payoff accruing from time $T+1$ onward. Note that, whatever t tells his successor, the actions from $T+1$ onward will all be arbitrarily close to either $-\beta$ or $1-\beta$, according to whether the outcome of the review phase will be declared to be zero or one.

Now consider what the belief of player t would be at the beginning of period $T+1$, if he could observe the signal realizations in all periods from t to T. Assume that all players, including t, obey the equilibrium strategies we have described. Using Bayes' rule, this belief would clearly be p or more for those realizations of signals that entail an outcome of the review phase equal to one, and would be $1-p$ or less for those realizations of signals that entail an outcome of the review phase equal to zero. If he plays according to his equilibrium strategy, from $T+1$ onwards player t's payoff will be the one corresponding to an action close to $1-\beta$ in the former case and the one corresponding to an action close to $-\beta$ in the latter case. It then follows from the first part of the argument that, if t plays according to his equilibrium strategy, the continuation actions from $T+1$ onwards will be matched to what he prefers, contingent on every possible realizations of signals.

If on the other hand t deviates from his putative equilibrium strategy, it is not hard to see that the deviation will make his beliefs and preferred continuation actions from $T+1$ onwards mismatched with positive probability. Since δ is appropriately high, gains accruing before $T+1$ can be ignored. Hence, no player $t \leq T$ wants to deviate from the proposed equilibrium.

Our next concern is limit learning in the public communication case.

7.2. Public Communication

Provided that the signal quality is not too high, limit learning cannot occur when communication is public. As is the case in Proposition 3, our next statement identifies a simple set of sufficient conditions.[19]

Proposition 4. *No Limit Learning with Public Communication: Fix any β and r and suppose that $p < \sqrt{2}/(1+\sqrt{2})$. Then limit learning is impossible in the model with public communication.*

As with other results, the formal proof is in the Appendix, and here we proceed with a brief sketch of the argument. We then compare the public to the private communication case to understand why exactly limit learning possible in one case but not the other.

Fix a discount factor δ and an equilibrium. From the proof of Proposition 2 we know that there is a greatest lower bound to the values that the belief of any player can take in this equilibrium. Denote this bound by $\underline{x}$.

The argument proceeds by contradiction. Suppose that the proposition is false, then as δ approaches one we must be able to find a sequence of equilibria with $\underline{x}$ approaching zero, so that limit learning can occur. (Limit learning of course requires more, but this is clearly a necessary condition for it to be possible at all.)

For simplicity of exposition, consider the case where the lower bound $\underline{x}$ is actually achieved. In this case there is some player, say t, who has a beginning-of-period belief x_t and who, after observing $s_t = 0$, chooses a mixed strategy that puts positive probability on a message giving player $t+1$ a belief of $\underline{x}$. Since the lower bound $\underline{x}$ is reached, all subsequent players will babble since informative messages would push beliefs below $\underline{x}$ when low signals are observed. Consequently, the action $\underline{x} - \beta$ is chosen in every future period.

Now suppose that t deviates by mimicking the type who observes $s_t = 1$. It is convenient to look at t's continuation payoff component by component. Fix a given period ahead, say $t+\tau$, and a sequence of realized signals between t and $t+\tau$. Two cases are possible. Given the sequence of signals and sequence of actions played with positive probability following the

[19] Note that even when the condition $p < \sqrt{2}/(1+\sqrt{2})$ is satisfied, Proposition 4 does not imply Proposition 2. This is simply because the latter shows the impossibility of full learning for *any* value of δ, while limit learning concerns values of δ approaching 1.

deviation, either the lower bound has been achieved at date $t+\tau$ and hence action $\underline{x}-\beta$ is chosen by player $t+\tau$, or the bound has not been achieved and an action $\hat{x}-\beta > \underline{x}-\beta$ is chosen instead. In the former case, clearly player t is indifferent in his $t+\tau$ payoff between following his equilibrium strategy and the hypothesized deviation.

Consider then the case in which action $\hat{x}-\beta > \underline{x}-\beta$ is chosen at $t+\tau$. Notice that since player $t+\tau$ does not know that t deviated, his beginning-of-period belief $\hat{x}$ is above the belief he would have held if he had known about player t's deviation. To be more precise, if player $t+\tau$ *had* known that t deviated, his belief, call it $\hat{x}^0$, would be located between one notch and two notches down from $\hat{x}$, the precise location depending on the mixed message strategy chosen by the type $s_t = 0$. This would also be t's belief at date $t+\tau$ (following the same realization of signals and mixed strategies as those observed by $t+\tau$). Consequently, player t can evaluate his payoff gain/loss in period $t+\tau$ from the deviation by comparing action $\hat{x}-\beta$ (following the deviation) to action $\underline{x}-\beta$ (if no deviation), and evaluating the payoffs using his belief $\hat{x}^0$.

The condition that $p < \sqrt{2}/(1+\sqrt{2})$ of the statement of the proposition now allows us to close the argument. In fact when p is low in this sense, so that each notch of updating makes an appropriately small difference to beliefs, straightforward algebra is sufficient to show the following. If $\underline{x}$ is sufficiently low, then the deviator, player t, strictly prefers action $\hat{x}-\beta$ taken by player $t+\tau$ following t's deviation to the action $\underline{x}-\beta$ that $t+\tau$ *would have taken* if no deviation occurred.

Given the arbitrary selection of the future player $t+\tau$, this implies that for $\underline{x}$ sufficiently low, player t finds the deviation hypothesized above strictly profitable, independently of the value of δ. This destroys the putative equilibrium and, hence, closes the argument.

It is worth asking: why is there a difference between private and public communication for limit learning? To address this, it is useful to revisit the construction in the proof outline of Proposition 3. There, a T period review phase determines what players should do based on a simple majority count of signals. Critically, once the majority threshold is reached, no new information is added. Players after date T do not know anything about the actual signal count other than the fact that a majority of one type of signal was reached. This creates a large enough discontinuity in payoffs between the 0 and 1 message so as to prevent a deviation. This coarseness of information is, in fact, essential to the construction. By contrast with public communication, the exact count is public. This destroys the putative construction above since certain paths of signals will differ substantially from their expected values. Hence, if along such paths, play were to proceed as if under the private communication review phase, then eventually a player would find it profitable to withhold his information thereby "free riding" on the information provision of others.

Finally, given the result it is instructive to piece together what we know from Propositions 3 and 4 about limit learning in the private and public communication cases. While the characterization is by no means complete because of the parametric restrictions imposed in Propositions 3 and 4, on a subset of the parameter space these two results allow for a comparison.

In particular, directly from the sufficient conditions stated in Propositions 3 and 4 we get that if $\beta < p - 1/2 < \sqrt{2}/(1+\sqrt{2}) - 1/2$, then limit learning is possible in the private communication case but impossible if communication is instead public.[20] This observation substantiates the remark we made in the introduction that the investigation of limit learning yields the insight that "more" (public) communication does not necessarily enable the organization to "learn more."

8. Mediated Communication

We now turn to the well-known theoretical benchmark case of mediated communication model (Forges, 1986, Myerson, 1982). In this case our main results are overturned: full learning can occur. But because mediated communication seems so unnatural in this context, in our view this strengthens our original no-full learning results.

Full learning equilibria under mediated communication are natural variations on those that yielded limit learning in Proposition 3 above, exploiting the ability to filter information to make appropriate use of the Weak Law of Large Numbers. In this sense, the mediated communication case sheds light on the comparison between Propositions 3 and 4 above. The ability to pass on filtered information down the line is (in principle) unlimited in the mediated communication case. In the case of private communication filtering is restricted by the fact that communication between two decision makers can only be mediated by the intervening players. In the of public communication case, no filtering is possible at all.

In the mediated communication model, we take it to be the case that each player t reports to a mediator the signal s_t that he observes, so that we can simply choose $M_t = \{0, 1\}$ for every t. The mediator then recommends an action to each player t on the basis of the reports of all players from 1 to $t-1$.

A mediated communication protocol $\mathcal{M}$ is a sequence of maps $\{\mu_t\}_{t=0}^{\infty}$. Each μ_t is from

[20] $\sqrt{2}/(1+\sqrt{2}) - 1/2 \approx 0.086$.

$\{0,1\}^{t-1}$ into $\mathbb{R}$. The interpretation is that $\mu_{t-1}(m_1,\ldots,m_{t-1})$ is the action that the mediator recommends to player t if the reports of the previous players are $(m_1,\ldots,m_{t-1})$. Consistently with the notation we established in Subsection 3.1, we set $m^{t-1} \equiv \mu_{t-1}(m_1,\ldots,m_{t-1})$.

We call $\mathcal{M}^*$ a mediated communication equilibrium if it satisfies the standard conditions that, given $\mathcal{M}^*$, all players are willing to truthfully reveal the signal they observe, and all players are willing to take whatever action $a_t = m^{t-1}$ prescribed by $\mathcal{M}^*$. Notice that any given $\mathcal{M}^*$, mapping sequences of signals into actions, induces a stochastic process $\{a_t\}_{t=1}^{\infty}$.

Recall that, using (2) and (3), player t is willing to take action a_t if and only if his belief is equal to $x_t = a_t + \beta$. Therefore, in line with Definition 1, we say that a mediated communication equilibrium $\mathcal{M}^*$ induces full learning of the state if and only if the induced stochastic process $\{a_t + \beta\}_{t=1}^{\infty}$ converges in probability to the true state.

Our next result asserts that full learning can obtain in the mediated communication case.

Proposition 5. *Full Learning With a Mediator: Fix δ, p and r. Suppose that $\beta < p - 1/2$. Then there is a mediated communication equilibrium $\mathcal{M}^*$ that induces full learning of the state.*

The proof of Proposition 5 is omitted for reasons of space. Once one has the details of the proof of Proposition 3, it is largely a notational and reinterpretational exercise. As with our previous results, we provide a sketch of the argument here.

The communication equilibrium $\mathcal{M}^*$ that we propose induces full learning of the state using an infinite sequence of review phases like the one in the proof of Proposition 3, increasing in length through time. For ease of exposition we will refer to the mediator recommending an action to a player and to the belief induced by the mediator's message interchangeably. The former equals the latter minus β so that no ambiguity will ensue.

Fix a small number ε. Number the phases as 1 through to n and beyond, and let T_n be the length of each phase. Recall that δ, p and r are given, and that, by assumption, the condition $\beta < p - 1/2$ of Proposition 3 is also satisfied here.

All players in each review phase, say n, receive a message from the mediator recommending an action based on the outcome of review phase $n-1$ (the empty message in the case of review phase 1), and report truthfully to the mediator the signal they observe.

Each review phase n is sufficiently long (but may exceed the minimum length required for this as will be clear below) to ensure that, as in the proof of Proposition 3, via the Weak Law of Large Numbers, the belief of players in review phase $n+1$ are within ε^n of the true state with probability at least $1-\varepsilon^n$. This can be done by picking T_n large enough, just as in the proof of Proposition 3. For ease of reference below, call this the precision requirement for T_n. Notice that of course as n grows large and hence ε^n shrinks to zero this is sufficient to achieve full learning of the state as required.

We then need to check that no player wants to deviate from any of the prescriptions of $\mathcal{M}^*$ in any of the review phases. In one respect, this is simpler than in the case of Proposition 3. This is so because no player can affect the actions of other players in the same review phase. The mediator does not reveal to them any information pertaining to the current review phase. So, the incentives to reveal truthfully his signal for a player in review phase $n-1$ only depend on the actions of players in review phases n and beyond. It then follows that by picking T_n sufficiently large we can ensure that the incentives for players in review phase $n-1$ are determined by their (strict) preferences over actions that will be taken during review phase n, regardless of the fixed value of the discount factor δ. Call this the incentive requirement for T_n. We then pick each T_n to be large enough so that both the precision requirement and the incentive requirement are met, and, aside from one caveat, the argument is complete.

The caveat is that in our sketch of the proof of Proposition 3 we dealt with a single symmetric review phase that for simplicity was assumed to begin with a prior r equal to one half. Clearly, with multiple review phases, as n increases the staring point of each review phase will be further and further away from the symmetry guaranteed by $r = 1/2$. This point can be dealt in the same way as in the formal proof of Proposition 3 presented in the Appendix. The outcome of each review phase will not be determined just by simple majority, and the length requirements of T_n mentioned above will have to take these asymmetries into account.

9. Conclusions and Extensions

9.1. More General Payoffs

Working with a specific functional form for payoffs as in (2) ensures tractability and allows some key parametric comparisons between the private and public communications cases. However, it is clearly more restrictive than one would like, and hence begs the question of how far this can be relaxed. To fix ideas, imagine that we wrote a general form of (2) as

$$(1-\delta)\left[v(a_t,\omega)+\sum_{\tau=1}^{\infty}\delta^{\tau}u(a_{t+\tau},\omega)\right] \tag{6}$$

which of course reduces to (2) when $v(a_t, \omega) = -(\omega - a_t)^2 - 2\beta a_t$ and $u(a_{t+\tau}, \omega) = -(\omega - a_{t+\tau})^2$. The question we are pursuing in this subsection is then what forms of $v(\cdot,\cdot)$ and $u(\cdot,\cdot)$ can we use and ensure that our results are still viable.

Although the assumption embodied in (2) can be relaxed in both cases, the degree to which we know this to be true depends on whether we are considering the case of private or public communication. Everything we say here about this issue depends on the structure of the proofs of our claims as they currently are. It follows that further generalizations cannot be ruled out, but depend on insights that would have to stem from proofs that are different from the ones we presented above.

In the case of private communication, our arguments for Propositions 1 and 3 generalize in a fairly direct way to a very broad class of possible payoffs. A set of sufficient conditions are as follows. Both u and v in (6) are continuous and concave over the appropriate range, and the maximizing values of a_t for both u and v exist and are finite for both $\omega = 1$ and $\omega = 0$. The bias term works consistently in one direction so that, regardless of the state, the maximizing value of a_t for u must be larger than that for v. Finally, to dispose of some uninteresting cases we need the equivalent of assuming $\beta < 1/2$ when the payoffs are as in (2).[21] To this end it is enough to require that the value of u given $\omega = 0$ is larger if we plug in the value of a_t that maximizes v given $\omega = 0$ than if we use the a_t that maximizes v given $\omega = 1$.[22]

The conditions we know to be sufficient for our arguments in the case of public communication to be viable for more general payoffs are more stringent and less transparent than in the case of private communication. As before, we require both u and v in (6) to be continuous and concave over the appropriate range, and the maximizing values of a_t for both u and v to exist finite for both $\omega = 1$ and $\omega = 0$. The bias term should be "separable" in the sense that for some $f(\cdot)$ we can write $u(a_t, \omega) = v(a_t, \omega) + f(a_t)$ for all a_t and $\omega \in \{0, 1\}$, with $f(\cdot)$ (weakly) increasing and convex. Finally, if we let $\mathbf{a}^v(x)$ be the maximizing value of a_t for $xv(a_t, 1) + (1 - x)v(a_t, 0)$, then $\mathbf{a}^v(x)$ should be (weakly) convex.

The conditions for the public communication case that we have described are clearly met in the linear-quadratic case in (2) and in a "neighborhood" (appropriately defined) of it. It is not hard to see that some other explicitly defined functional forms will satisfy them too.

[21] See footnote 8 above.

[22] The same conditions are sufficient to generalize the argument we use to prove Proposition 5.

9.2. Concluding Remarks and Further Extensions

We examine large, complex ongoing organizations in which information flows through players who are both senders and receivers of it. In particular, we analyze a model that fits an ongoing firm or other organization that is led by a sequence of imperfectly informed decision makers, with each decision maker getting information from his predecessors and transmitting it down the line to future decision makers. To take good decisions our players must, to borrow from Isaac Newton's famous quote, be "standing on the shoulders of giants.[23]" We analyze the limits to this process when preferences are not perfectly aligned and hence players act strategically in the transmission of information.

We examine both polar cases of private and public communication from one decision maker to his immediate successor in one case, and to all subsequent players in the other. The main results show that in either model, full learning is impossible.

Demonstrating the impossibility of full learning involves a number of difficulties in both cases. Even under a set of simplifying assumptions, in most cases a direct evaluation of continuation payoffs proves an intractable route to follow in our model. As a consequence our main results are proved using a variety of techniques to bound these payoffs appropriately.

Our proof techniques do not allow a strict characterization of the features of the equilibria of the model in addition to the no full learning claims. This is because of the bounding arguments we have just referred to. It is worth noticing that in the case of public communication we do know that in equilibrium there are "uniform" bounds that learning cannot exceed, and that periods of babbling and learning can alternate. In the case of private communication we know that babbling and learning can alternate and that there are equilibria in which all past is "erased" at exogenously given dates and behavior starts afresh as in period 0.

We also ask whether limit learning can be sustained in the sense of a sequence of equilibria that yields full learning in the limit, as discounting shrinks to zero. The two cases of private and public communication yield different results in this case. For some configurations of parameters, limit learning is possible with private communication but it is impossible in the public communication case.

Finally, we examined the theoretical benchmark case of mediated communication, and showed full learning can be sustained in this case.

[23]In a 1676 letter to Robert Hooke. A popularized account of the context for this exceedingly well known quote and of the relationship between Newton and Hooke can be found in the "Introduction" and the "Sir Isaac Newton" chapter of Hawking (2003).

Our results in this paper generalize to the case in which the state takes values in a finite set, and to correspondingly richer signal structures, provided that the imperfect signals are still informative and i.i.d.

An interesting case to consider in future work is that of a signal quality that can change through time. These changes could take place either exogenously because of, say, technological advances, or endogenously, because of investment choices on the part of the players.

Appendix

A.1. The Proof of Proposition 1

The proof proceeds by contradiction. If full learning is possible then there is a player with a sufficiently large beginning-of-period belief who sends informative messages. We show that this player prefers to conceal his information and send the message which is interpreted by his successor as the strongest evidence in favor of the state $\omega = 1$.

Before proceeding with the main line of argument, we argue that mixing in the original game is equivalent to allowing each player to observe a private randomization device and then use pure strategies. As a further preliminary, we show that it is without loss of generality to restrict attention to games in which all messages are sent with positive probability.

Given the original game with private communication, consider a "modified" game which is identical to the original one except that in each period $t = 1, 2, \ldots$ player t, after observing s_t and before choosing m_t, observes the realization of a random variable $\gamma_t \in [0, 1]$ (no-one else observes γ_t). This random variable plays the role of a randomization device and is distributed uniformly, independently of the state, the signals and all other $\gamma_{t'}$, $t' \neq t$.

With a minor abuse of notation, we let $\boldsymbol{\sigma} = \{\sigma_t\}_{t=1}^{\infty}$ denote a *pure* strategy profile of the modified game. For every t, $\sigma_t(m_{t-1}, s_t, \gamma_t)$ is the message sent by player t when he receives the message m_{t-1} and observes the realization s_t of the signal and the realization γ_t of the randomization device.

Definition A.1: *Fix any δ, p, β and r. Let $\boldsymbol{\sigma}^*$ be an arbitrary equilibrium strategy profile of either the original or modified game with message sets $M = \{M_t\}_{t=0}^{\infty}$. Let $\hat{\boldsymbol{\sigma}}^*$ be another arbitrary equilibrium strategy profile of either the original or modified game with message sets $\hat{M} = \{\hat{M}_t\}_{t=0}^{\infty}$. We say that $\boldsymbol{\sigma}^*$ and $\hat{\boldsymbol{\sigma}}^*$ are* outcome equivalent *if and only if they induce the same stochastic process $\{x_t\}_{t=1}^{\infty}$ on the players' posteriors for every realization of the state ω.*

Lemma A.1: *Fix any δ, p, β, r and any (pure or mixed) equilibrium strategy profile $\boldsymbol{\sigma}^*$ of the original game with messages $M = \{M_t\}_{t=0}^{\infty}$ .Then there exist an $\hat{M} = \{\hat{M}_t\}_{t=0}^{\infty}$ with $\hat{M}_t \subseteq M_t$ for every t and a $\hat{\boldsymbol{\sigma}}^*$ which is an equilibrium of the modified game in* pure *strategies with messages $\hat{M}$ and satisfies the following. Given $\hat{\boldsymbol{\sigma}}^*$ and $\hat{M}$, all messages are sent with positive probability regardless of the value of $\omega \in \{0, 1\}$, and $\hat{\boldsymbol{\sigma}}^*$ is outcome equivalent to $\boldsymbol{\sigma}^*$ according to Definition A.1.*

Proof: We only provide a sketch since the arguments involved are standard.

The randomization device γ_t does not provide player t with any information about the state or the behavior of the other players. When the player is indifferent (and is supposed to randomize in the original game), then he can use the randomization device to replicate his mixed strategy in the original game with a pure strategy in the modified game.

If the message spaces are modified to exclude those message that are not sent with positive probability in equilibrium, we simply reduce the set of possible deviations. Hence the new (outcome equivalent) strategy profile as in the statement of the lemma must be an equilibrium. ■

Remark A.1: *In view of Lemma A.1, we only consider specifications of M and pure strategy equilibria of the modified game $\boldsymbol{\sigma}^*$, in which all messages are sent with positive probability for both values of $\omega \in \{0,1\}$ throughout the rest of the proof of Proposition 1.*

Before proceeding further, some extra notation is needed. The entire batch is defined taking as given an M and an equilibrium strategy profile $\boldsymbol{\sigma}^*$ as in Remark A.1.

To begin with, define

$$X_t^B = \bigcup_{m_{t-1} \in M_{t-1}} x_t(m_{t-1}) \tag{A.1}$$

so that X_t^B is the set of all possible beginning-of-period beliefs of player t that $\omega = 1$. Because of Remark A.1 these can all be computed using Bayes' rule.

We let $x_t^E(m_{t-1}, s_t, \gamma_t)$ denote the end-of-period belief of player t that $\omega = 1$ when he receives message m_{t-1} and observes signal $s_t = 0, 1$ and γ_t during period t. Notice that using Bayes rule

$$x_t^E(m_{t-1}, s_t, \gamma_t) = \frac{x_t(m_{t-1})\Pr(s_t|\omega = 1)}{x_t(m_{t-1})\Pr(s_t|\omega = 1) + [1 - x_t(m_{t-1})]\Pr(s_t|\omega = 0)}$$

for every $\gamma_t \in [0,1]$. Although the end-of-period belief does not vary with the realization of the randomization device, we find it useful to condition the belief on everything the player knows.

Consider player t and any message $m_{t-1} \in M_{t-1}$ that player t can receive. We let $U_t(m_{t-1}, \omega)$ be the discounted continuation payoff of player t if the state is ω and he behaves as if he has received message m_{t-1} (in this case, we say player t *behaves according to* m_{t-1}). The payoff is computed after player t has chosen his action and before he observes his signal s_t and his randomization device γ_t. The continuation payoff is normalized so that weight $(1-\delta)$ is given to the expected payoff in period $t+1$ and weight $(1-\delta)\delta^\tau$ is given to the expected payoff in each period $t + \tau + 1$.

Formally, for any $\tau \geq 1$, let $\mathrm{a}_{t+\tau}(m_{t-1}, s^{t,t+\tau-1}, \gamma^{t,t+\tau-1})$ denote the action that player $t+\tau$ will choose if player t behaves according to m_{t-1} and the sequences of signals and randomization devices between period t and period $t+\tau-1$ are $s^{t,t+\tau-1} = (s_t, s_{t+1}, \ldots, s_{t+\tau-1})$ and $\gamma^{t,t+\tau-1} = (\gamma_t, \gamma_{t+1}, \ldots, \gamma_{t+\tau-1})$, respectively.[24]

Then $U_t(m_{t-1}, \omega)$ is given by

$$U_t(m_{t-1}, \omega) =$$

$$-(1-\delta) \sum_{\tau=1}^{\infty} \delta^{\tau-1} \sum_{s^{t,t+\tau-1} \in \{0,1\}^\tau} \int_{\gamma^{t,t+\tau-1} \in [0,1]^\tau} \Pr(s^{t,t+\tau-1}|\omega) \tag{A.2}$$

$$[\omega - \mathrm{a}_{t+\tau}(m_{t-1}, s^{t,t+\tau-1}, \gamma^{t,t+\tau-1})]^2 \; \mathrm{d}\,\gamma^{t,t+\tau-1}$$

[24]Throughout the rest, for any $t' \geq t$, the arrays $(s^t, \ldots, s^{t'})$ and $(\gamma^t, \ldots, \gamma^{t'})$ will be denoted by $s^{t,t'}$ and $\gamma^{t,t'}$, respectively.

Lastly, notice that $U_t(m_{t-1}, \omega)$ as defined in (A.2) satisfies the recursive relationship

$$
\begin{aligned}
&U_t(m_{t-1}, \omega) = \\
&\Pr(s_t = 0|\omega) \int_{\gamma_t \in [0,1]} \left\{ -(1-\delta)[\omega - \mathrm{a}_{t+1}(m_{t-1}, s_t = 0, \gamma_t)]^2 + \delta U_{t+1}[\sigma_t(m_{t-1}, s_t = 0, \gamma_t), \omega] \right\} \, \mathrm{d}\,\gamma_t + \\
&\Pr(s_t = 1|\omega) \int_{\gamma_t \in [0,1]} \{ -(1-\delta)[\omega - \mathrm{a}_{t+1}(m_{t-1}, s_t = 1, \gamma_t)]^2 + \delta U_{t+1}[\sigma_t(m_{t-1}, s_t = 1, \gamma_t), \omega] \} \, \mathrm{d}\,\gamma_t
\end{aligned}
\tag{A.3}
$$

Lemma A.2: *Fix an M and $\boldsymbol{\sigma}^*$ as in Remark A.1. For any $\eta > 0$ there exists an $\varepsilon > 0$ such that the following is true for every t. Suppose that m_{t-1} and m'_{t-1} are two messages in M_{t-1} with $\min\{x_t(m_{t-1}), x_t(m'_{t-1})\} > 1 - \varepsilon$ and $x_t(m_{t-1}) \neq x_t(m'_{t-1})$. Then*

$$
|U_t(m_{t-1}, 1) - U_t(m'_{t-1}, 1) - U_t(m_{t-1}, 0) + U_t(m'_{t-1}, 0)| < \eta \tag{A.4}
$$

Proof: Omitted. Available at `http://www.anderlini.net/learning-omitted-proofs.pdf`

We are now ready to proceed with the main part of the argument for Proposition 1.

Proof of Proposition 1: Fix any δ, p, β and r. Let also an M and $\boldsymbol{\sigma}^*$ as in Remark A.1. Suppose by way of contradiction that the equilibrium $\boldsymbol{\sigma}^*$ induces full learning as in Definition 1. Distinguish between the following two possibilities.

Case 1: For every $\varepsilon > 0$, we can find a $\bar{t}$ and a pair of messages $m_{\bar{t}-1}$ and $m'_{\bar{t}-1}$ with $\min\{x_{\bar{t}}(m_{\bar{t}-1}), x_{\bar{t}}(m'_{\bar{t}-1})\} > 1 - \varepsilon$ and $x_{\bar{t}}(m_{\bar{t}-1}) \neq x_{\bar{t}}(m'_{\bar{t}-1})$.

and

Case 2: There exists $\tilde{\varepsilon}$ such that for every t there is at most one element of X_t^B (the set of all possible beginning-of-period equilibrium beliefs of player t as in (A.1)) strictly greater than $1 - \tilde{\varepsilon}$.

Clearly one (and only one) of Cases 1 and 2 must hold.

We begin with Case 1. Let $\eta = (1-\delta)\beta/2\delta$. Using Lemma A.2, we can pick an $\varepsilon \in (0, \beta/2)$, a $\bar{t}$ and a pair of messages $m_{\bar{t}-1}$ and $m'_{\bar{t}-1}$ as specified in Case 1 such that

$$
|U_{\bar{t}}(m_{\bar{t}-1}, 1) - U_{\bar{t}}(m'_{\bar{t}-1}, 1) - U_{\bar{t}}(m_{\bar{t}-1}, 0) + U_{\bar{t}}(m'_{\bar{t}-1}, 0)| < \eta = \frac{(1-\delta)\beta}{2\delta} \tag{A.5}
$$

To keep notation usage down, during this part of the argument we let $m_{\bar{t}-1} = m$, $m'_{\bar{t}-1} = m'$, $x_{\bar{t}}(m_{\bar{t}-1}) = x$ and $x_{\bar{t}}(m'_{\bar{t}-1}) = x'$. Without loss of generality, assume $x > x'$.

Clearly, there exists a type $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1}, \hat{\gamma}_{\bar{t}-1})$ of player $\bar{t}-1$ who has the end-of-period belief $x_t^E(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1}, \hat{\gamma}_{\bar{t}-1}) \geq x'$ and sends message m'. We now show that type $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1}, \hat{\gamma}_{\bar{t}-1})$ has an incentive to deviate and send message m. To keep notation usage down again, during this part of the argument we also let $\hat{x} = x_t^E(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1}, \hat{\gamma}_{\bar{t}-1})$.

Let $\varphi(\hat{x})$ denote the difference between the continuation payoff of type $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1}, \hat{\gamma}_{\bar{t}-1})$ after he sends message m and the continuation payoff after he sends the equilibrium message m'. Using (2) and (A.3), the quantity $\varphi(\hat{x})$ can be written as

$$
\varphi(\hat{x}) = (1-\delta)\bar{\varphi}(\hat{x}) + \delta(A\hat{x} + B) \tag{A.6}
$$

where

$$\bar{\varphi}(\hat{x}) = -\hat{x}(1-x+\beta)^2 - (1-\hat{x})(x-\beta)^2 + \hat{x}(1-x'+\beta)^2 + (1-\hat{x})(x'-\beta)^2$$

and

$$A = U_{\bar{t}}(m,1) - U_{\bar{t}}(m',1) + U_{\bar{t}}(m',0) - U_{\bar{t}}(m,0) \qquad \text{and} \qquad B = -U_{\bar{t}}(m',0) + U_{\bar{t}}(m,0) \tag{A.7}$$

Notice that it follows from the incentive constraints of player $\bar{t}$ (to whom the second term of the payoff difference (A.6) applies) that

$$Ax + B \geq 0 \geq Ax' + B \tag{A.8}$$

which, since $x > x'$, implies that $A \geq 0$.

We now show that $\varphi(\hat{x})$ is strictly positive and, therefore that type $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1}, \hat{\gamma}_{\bar{t}-1})$ of player $\bar{t}-1$ has an incentive to deviate from his equilibrium strategy. We need to distinguish between two cases depending on whether $\hat{x}$ is smaller or greater than x. We start with the case $\hat{x} \in [x', x)$.

A lower bound for the expression $\bar{\varphi}(\hat{x})$ can be computed in the following way

$$\bar{\varphi}(\hat{x}) = (x-x')(2\beta + 2\hat{x} - x - x') > (x-x')2(\beta - \varepsilon) > (x-\hat{x})\beta \tag{A.9}$$

where the first inequality follows from $\hat{x} > 1-\varepsilon$ (and $x < 1$ and $x' < 1$), while the second inequality follows from $\varepsilon < \beta/2$.

Consider now again the entire payoff difference $\varphi(\hat{x})$. We have

$$\begin{aligned}\varphi(\hat{x}) > (1-\delta)(x-\hat{x})\beta + \delta(A\hat{x}+B) = \\ (1-\delta)(x-\hat{x})\beta - \delta(x-\hat{x})A + \delta(Ax+B) \geq \\ (x-\hat{x})[(1-\delta)\beta - \delta A] > (x-\hat{x})(1-\delta)\frac{\beta}{2} > 0\end{aligned} \tag{A.10}$$

where the first inequality comes from (A.6) and (A.9), the second follows from (A.8), the third follows from (A.5) and the definition of A in (A.7), and the last one from the fact that we are considering the case in which $\hat{x} \in [x', x)$.

Consider now the case $\hat{x} \geq x$. We have that (recall that $A \geq 0$)

$$\begin{aligned}\varphi(\hat{x}) = (1-\delta)(x-x')(2\beta + 2\hat{x} - x - x') + \delta(A\hat{x}+B) \geq \\ (1-\delta)(x-x')(2\beta + x - x') + \delta(Ax+B) > 0\end{aligned}$$

This proves that type $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1}, \hat{\gamma}_{\bar{t}-1})$ of player $\bar{t}-1$ has an incentive to deviate from the equilibrium prescription of sending message m', and hence concludes the argument for Case 1.

We now turn to Case 2. By our contradiction hypothesis, $\boldsymbol{\sigma}^*$ induces full learning of the state. Therefore, since we are in Case 2 as identified above, we can find $\tilde{t}$ such that for every $t > \tilde{t}$ there exists a unique beginning-of-period belief for player t above $1-\tilde{\varepsilon}$. This belief must be the largest element in X_t^B and we denote it by $x_t^{\max}$.

Next, fix an ε' such that

$$\frac{(1-\tilde{\varepsilon})p}{(1-\tilde{\varepsilon})p + \tilde{\varepsilon}(1-p)} < 1 - \varepsilon' \tag{A.11}$$

Given ε' we can find a period $t(\varepsilon') > \tilde{t}$ such that for every $t > t(\varepsilon')$

$$\frac{x_t^{\max}(1-p)}{x_t^{\max}(1-p) + (1 - x_t^{\max})p} > 1 - \varepsilon' \tag{A.12}$$

For every t let $M_{t-1}^{\max} \subset M_{t-1}$ be the set of messages that induce belief $x_t^{\max}$, so that

$$M_{t-1}^{\max} = \{m_{t-1} \in M_{t-1} \mid x_t(m_{t-1}) = x_t^{\max}\}$$

Because $\boldsymbol{\sigma}^*$ induces full learning, the sequence $\{x_t^{\max}\}_{t=1}^{\infty}$ must converge to one. Using (A.11) and (A.12) it is then easy to show that there must exist a period $\bar{t} - 1 > t(\varepsilon')$ and two types $(m_{\bar{t}-2}, s_{\bar{t}-1}, \gamma_{\bar{t}-1})$ and $(m'_{\bar{t}-2}, s'_{\bar{t}-1}, \gamma'_{\bar{t}-1})$ in $M_{\bar{t}-2}^{\max} \times \{0,1\} \times [0,1]$ which send two distinct messages, say m and m' respectively, such that

$$x_{\bar{t}}(m) > 1 - \varepsilon' \qquad \text{and} \qquad x_{\bar{t}}(m') < 1 - \tilde{\varepsilon}$$

It is also easy to check that when ε' is sufficiently small type $(m'_{\bar{t}-2}, s'_{\bar{t}-1}, \gamma'_{\bar{t}-1})$ has an incentive to deviate and send message m instead of m' as the equilibrium prescribes. The analysis is very similar to the analysis of Case 1 and we omit the details. The argument in Case 2 is therefore concluded. Hence, the proof of Proposition 1 is now complete. ■

A.2. The Proof of Proposition 2

As with Proposition 1, the proof proceeds by contradiction. We first show that if full learning obtains then there must be a player with a sufficiently large beginning-of-period belief who sends a message that fully reveals a certain realization of the signal (i.e., the message is sent with positive probability if and only if the player observes that realization). This creates a discrete gap in the beliefs that different equilibrium messages of this player induce among his successors.

The next and final part of the argument is to show that the presence of this gap implies that this player has a strict incentive to send the message that induces the largest belief in his successor, independently of what he observes. This contradicts the full revelation claim that we established in the first place, and hence suffices to close the proof.

We formalize the argument for the case $M_t = \{0,1\}$, for every t ($M_0 = \emptyset$). Generalizing to any collection of finite message sets is straightforward and we omit the details.

Recall that in the public communication case player t observes $m^{t-1} = (m^0, m_1, \ldots, m_{t-1})$, the sequence of all messages sent by all players up to and including $t-1$ (set $m_0 = m^0 = \emptyset$). Recall also that, with a minor abuse of notation, we write $\sigma_t(m^{t-1}, s_t) = m_t$ if player t sends message m_t with probability one.

Without loss of generality, throughout the argument we consider strategies such that for every m^{t-1}, $x_{t+1}\left(m^{t-1}, m_t = 1\right) \geqslant x_{t+1}\left(m^{t-1}, m_t = 0\right)$.

Fix a strategy profile $\boldsymbol{\sigma}$. We then let $\hat{U}_t(m^t, \omega)$ denote the discounted continuation payoff of player t if the state is ω and player $t+1$ and all his successors observe the sequence of messages m^t in the first t periods. The payoff is computed after player t chooses his action.

Lemma A.3: *Suppose that $\boldsymbol{\sigma}$ is an equilibrium strategy profile and that for some m^{t-1},* $\text{Supp}\ \sigma_t(m^{t-1},\, s_t = 0) = \text{Supp}\ \sigma_t(m^{t-1},\, s_t = 1) = \{0,1\}$. *Then for every $\omega \in \{0,1\}$, $\hat{U}_t(m^{t-1}, m_t = 0, \omega) = \hat{U}_t(m^{t-1}, m_t = 1, \omega)$.*

Proof: The claim is a direct consequence of the fact that player t's expected payoff from sending a given message m_t is a linear combination of $\hat{U}_t(m^{t-1}, m_t, \omega = 0)$ and $\hat{U}_t(m^{t-1}, m_t, \omega = 1)$ with weights given by t's beliefs about ω. The rest of the details are omitted. ■

Definition A.2: *Let $\boldsymbol{\sigma}$ be an equilibrium strategy profile. We say that $\boldsymbol{\sigma}$ has beliefs bounded above if there exists $\varepsilon > 0$ such that for every t and every m^{t-1} we have that $x_t(m^{t-1}) \leq 1 - \varepsilon$. Symmetrically, we say that $\boldsymbol{\sigma}$ has beliefs bounded below if there exists $\varepsilon > 0$ such that for every t and every m^{t-1} we have that $\varepsilon \leq x_t(m^{t-1})$.*

Lemma A.4: *Suppose that $\boldsymbol{\sigma}$ is an equilibrium strategy profile that does not have beliefs that are bounded above. Then for every $\varepsilon > 0$, we can find a t and an m^{t-1} such that $x_t(m^{t-1}) > 1-\varepsilon$ and $\operatorname{Supp} \sigma_t(m^{t-1}, s_t = 0) \neq \operatorname{Supp} \sigma_t(m^{t-1}, s_t = 1)$.*

Proof: Let $\boldsymbol{\sigma}$ be an equilibrium strategy profile that does not have beliefs that are bounded above. Notice that this directly implies that for every $\varepsilon > 0$, there exist t and m^{t-1} such that $x_t(m^{t-1}) > 1 - \varepsilon$ and $\sigma_t(m^{t-1}, s_t = 0) \neq \sigma_t(m^{t-1}, s_t = 1)$.

Suppose now, by contradiction, that there exists $\varepsilon' > 0$ such that for all t and m^{t-1} that satisfy $x_t(m^{t-1}) \geq 1 - \varepsilon'$, we have that $\operatorname{Supp} \sigma_t(m^{t-1}, s_t = 0) = \operatorname{Supp} \sigma_t(m^{t-1}, s_t = 1)$.

Choose $\bar{t}$ and $m^{\bar{t}-1}$ such that

$$\sigma_{\bar{t}}(m^{\bar{t}-1}, s_{\bar{t}} = 0) \;\neq\; \sigma_{\bar{t}}(m^{\bar{t}-1}, s_{\bar{t}} = 1) \tag{A.13}$$

and

$$x_{\bar{t}}(m^{\bar{t}-1}) \;>\; \frac{(1-\varepsilon')p}{(1-\varepsilon')p + \varepsilon'(1-p)} \tag{A.14}$$

Because of (A.13) in equilibrium, the two pairs $(m^{\bar{t}-1}, m_t = 0)$ and $(m^{\bar{t}-1}, m_t = 1)$ are followed by two stochastic processes over the messages and the actions in periods $\bar{t}+1, \bar{t}+2, \ldots$. We let $\hat{\mathbf{m}}^k$ and $\hat{\mathbf{a}}^k$, $k \in \{0,1\}$ denote the stochastic process, over the messages and the actions respectively, that, in equilibrium, follow the pair $(m^{\bar{t}-1}, m_t = k)$.

We now construct two (deterministic) sequences of message $\mathbf{m}^0 = \{m_t^0\}_{t=\bar{t}}^{\infty}$ and $\mathbf{m}^1 = \{m_t^1\}_{t=\bar{t}}^{\infty}$, that are related to the two stochastic processes we have defined, as follows. Both sequences will be constructed by induction. We begin with $\mathbf{m}^0$.

Let $m_{\bar{t}}^0 = 0$, fix a $t \geq 1$ and assume $m_{\bar{t}}^0, \ldots, m_{\bar{t}+t-1}^0$ have been set. We then describe the inductive step that yields the value of $m_{\bar{t}+t}^0$. We distinguish between two cases.

Case 1: Suppose that $x_{\bar{t}+t}(m^{\bar{t}-1}, m_{\bar{t}}^0, \ldots, m_{\bar{t}+t-1}^0) \geq 1 - \varepsilon'$.[25] If $\sigma_{\bar{t}+t}(m^{\bar{t}-1}, m_{\bar{t}}^0, \ldots, m_{\bar{t}+t-1}^0, s_{\bar{t}+t} = 0) = \sigma_{\bar{t}+t}(m^{\bar{t}-1}, m_{\bar{t}}^0, \ldots, m_{\bar{t}+t-1}^0, s_{\bar{t}+t} = 1) = 1$, then set $m_{\bar{t}+t}^0 = 1$. In all other cases, set $m_{\bar{t}+t}^0 = 0$.

Case 2: Suppose that $x_{\bar{t}+t}(m^{\bar{t}-1}, m_{\bar{t}}^0, \ldots, m_{\bar{t}+t-1}^0) < 1 - \varepsilon'$. If $\sigma_{\bar{t}+t}(m^{\bar{t}-1}, m_{\bar{t}}^0, \ldots, m_{\bar{t}+t-1}^0, s_{\bar{t}+t} = 0) = \sigma_{\bar{t}+t}(m^{\bar{t}-1}, m_{\bar{t}}^0, \ldots, m_{\bar{t}+t-1}^0, s_{\bar{t}+t} = 1) = 0$, then set $m_{\bar{t}+t}^0 = 0$. In all other cases, set $m_{\bar{t}+t}^0 = 1$.

We now turn to the sequence $\mathbf{m}^1$. Symmetrically to what we did for $\mathbf{m}^0$, let $m_{\bar{t}}^0 = 1$, fix a $t \geq 1$ and assume $m_{\bar{t}}^1, \ldots, m_{\bar{t}+t-1}^1$ have been set. We then describe the inductive step that yields the value of $m_{\bar{t}+t}^t$.

[25] The notation $(m^{\bar{t}-1}, m_{\bar{t}}^0, \ldots, m_{\bar{t}+t-1}^0)$ obviously denotes the sequence of messages of length $\bar{t}+t$ in which the first $\bar{t}$ messages are equal to $m^{\bar{t}-1}$ and the remaining messages are equal to $m_{\bar{t}}^0, \ldots, m_{\bar{t}+t-1}^0$.

Set $m^1_{\bar{t}+t} = 0$ if $\sigma_{\bar{t}+t}(m^{\bar{t}-1}, m^1_{\bar{t}}, \ldots, m^1_{\bar{t}+t-1}, s_{\bar{t}+t} = 0) = \sigma_{\bar{t}+t}(m^{\bar{t}-1}, m^1_{\bar{t}}, \ldots, m^1_{\bar{t}+t-1}, s_{\bar{t}+t} = 1) = 0$.[26] In all other cases, set $m^1_{\bar{t}+t} = 1$.

Given the equilibrium $\boldsymbol{\sigma}$ and the initial messages $m^{\bar{t}-1}$, the two sequences of messages $\mathbf{m}^0$ and $\mathbf{m}^1$ we have just constructed induce two sequences of actions, $\mathbf{a}^0 = \{a^0_t\}_{t=\bar{t}}^{\infty}$ and $\mathbf{a}^1 = \{a^1_t\}_{t=\bar{t}}^{\infty}$, respectively. Using (3) we have $a^k_t = x_t\left(m^{\bar{t}-1}, m^k_{\bar{t}}, \ldots, m^k_{\bar{t}+t-1}\right) - \beta$ for any $k \in \{0, 1\}$ and $t \geq 1$.

Next, note that it follows from (A.13) and (A.14) that

$$\operatorname{Supp} \sigma_{\bar{t}}(m^{\bar{t}-1}, s_{\bar{t}} = 0) = \operatorname{Supp} \sigma_{\bar{t}}(m^{\bar{t}-1}, s_{\bar{t}} = 1) = \{0, 1\} \tag{A.15}$$

Consider now a fictitious incarnation of player $\bar{t} - 1$. This is a version of player $\bar{t} - 1$ (with the same quadratic preferences of the real player $\bar{t} - 1$) at the end of period $\bar{t} - 1$, who has somehow been endowed with the knowledge that $\omega = 1$. Call this fictitious agent, player $<\bar{t} - 1, \omega = 1>$. Using (A.15) and Lemma A.3 we can see that player $<\bar{t} - 1, \omega = 1>$, is indifferent between the processes $\hat{\mathbf{m}}^0$ and $\hat{\mathbf{m}}^1$.

It is easy to see that player $<\bar{t} - 1, \omega = 1>$ is also indifferent between $\hat{\mathbf{m}}^1$ and $\mathbf{m}^1$. This is because the sequence $\mathbf{m}^1$ is constructed by replacing the random message of the process $\hat{\mathbf{m}}^1$ with the message 1 in all the cases in which the (future) players are indifferent between message 0 and message 1, regardless of the signal they observe.

In a similar way, it can be seen that player $<\bar{t} - 1, \omega = 1>$, weakly prefers the sequence $\mathbf{m}^0$ to the stochastic process $\hat{\mathbf{m}}^0$. This implies that for player $<\bar{t} - 1, \omega = 1>$ the sequence $\mathbf{m}^0$ is weakly better than the sequence $\mathbf{m}^1$.

However, for each $t = \bar{t} + 1, \bar{t} + 2, \ldots$ we have

$$a^1_t > x_{\bar{t}}(m^{\bar{t}-1}) - \beta > a^0_t \tag{A.16}$$

Clearly, (A.16) implies that player$<\bar{t} - 1, \omega = 1>$, must strictly prefer the sequence $\mathbf{m}^1$ to the sequence $\mathbf{m}^0$. Therefore, we have reached a contradiction. Hence, the proof of the lemma is now complete. ■

Remark A.2: *Suppose that $\boldsymbol{\sigma}$ is an equilibrium strategy profile that does not have beliefs that are bounded above. It follows from Lemma A.4 that at least one of the following two possibilities must be true.*

Case 1: For every $\varepsilon > 0$, there exist t and m^{t-1} such that $x_t(m^{t-1}) > 1 - \varepsilon$ and

$$\sigma_t(m_t = 0 | m^{t-1}, s_t = 0) = 1, \quad \sigma_t(m_t = 1 | m^{t-1}, s_t = 1) > 0 \tag{A.17}$$

Case 2: For every $\varepsilon > 0$, there exist t and m^{t-1} such that $x_t(m^{t-1}) > 1 - \varepsilon$ and

$$\sigma_t(m_t = 0 | m^{t-1}, s_t = 0) > 0, \quad \sigma_t(m_t = 1 | m^{t-1}, s_t = 1) = 1 \tag{A.18}$$

Proof of Proposition 2: It obviously suffices to prove that any equilibrium strategy profile $\boldsymbol{\sigma}$ must have beliefs that are bounded above (in fact this is obviously a stronger statement than needed). Before proceeding further we also remark that our entire line of proof could be replicated, mutatis mutandis, to show that equilibrium beliefs must also be bounded below.

To prove that beliefs are bounded above, we proceed by contradiction, and postulate a putative equilibrium strategy profile $\boldsymbol{\sigma}^*$ that does not have beliefs that are bounded above.

[26] See footnote 25.

Given such a $\boldsymbol{\sigma}^{\boldsymbol{*}}$, from Remark A.2, one (or both) of Cases 1 and 2 must hold. In the remainder of the argument, we assume that Case 1 holds. The argument for Case 2 is completely analogous and hence omitted.

Since Case 1 holds we can find a player t and an m^{t-1} with t's posterior after observing $s_t = 0$ (that is $x_t^E(m^{t-1}, s_t = 0)$) arbitrarily close to 1, who will play $m_t = 0$ with probability 1. To close the proof of Proposition 2 we now proceed to show he prefers to deviate and send message $m_t = 1$.

Consider then such a player t who observes m^{t-1}. From (A.17) and Bayes' rule, the beliefs of player $t+1$ at the beginning of $t+1$, $x_{t+1}(m^{t-1}, m_t)$, after observing the messages $m_t = 0$ and $m_t = 1$ respectively satisfy

$$\begin{aligned} x_t^E(m^{t-1}, s_t = 0) \leq x_{t+1}(m^{t-1}, m_t = 0) < z(x_t^E(m^{t-1}, s_t = 0)) \ < \\ x_{t+1}(m^{t-1}, m_t = 1) = h(x_t^E(m^{t-1}, s_t = 0)) \end{aligned} \tag{A.19}$$

where the functions z and h are defined by

$$z(x) = \frac{xp}{xp + (1-x)(1-p)} \qquad \text{and} \qquad h(x) = \frac{xp^2}{xp^2 + (1-x)(1-p)^2} \tag{A.20}$$

Consider a fictitious player, call him $<\!-1, x\!>$ who looks at our model from the end of a fictitious period "-1," just before the beginning of period 0, and assigns probability x to $\omega = 1$. Player $<\!-1, x\!>$ does not take any action and receives discounted quadratic loss payoffs from the actions of all future players just like all other players in the game. Crucially, player $<\!-1, x\!>$ also has the capacity to start the game endowing player $t = 0$ with a belief y that $\omega = 1$, which may or may not equal his own belief x.

Fix an arbitrary strategy profile $\boldsymbol{\sigma}$, which may or may not be an equilibrium strategy profile. Assume that play and the updating of beliefs proceeds according to $\boldsymbol{\sigma}$, with player 0 starting off with a (common knowledge among the following players) prior belief that $\omega = 1$ with probability y. We denote the payoff to player $<\!-1, x\!>$ by $V(x, y, \boldsymbol{\sigma})$.

Consider now again player t observing m^{t-1} as above. Using (A.19) his equilibrium continuation payoff from sending $m_t = 0$ after observing $s_t = 0$ is clearly bounded above by

$$\begin{aligned} &\max_{y,\boldsymbol{\sigma}} \quad V(x_t^E(m^{t-1}, s_t = 0), y, \boldsymbol{\sigma}) \\ &\text{s.t.} \quad x_t^E(m^{t-1}, s_t = 0) \leq y \leq z(x_t^E(m^{t-1}, s_t = 0)) \end{aligned} \tag{A.21}$$

Simlarly, using (A.19) again his equilibrium continuation payoff from sending $m_t = 1$ after observing $s_t = 0$ is clearly bounded below by

$$\min_{\boldsymbol{\sigma}} \quad V(x_t^E(m^{t-1}, s_t = 0), h(x_t^E(m^{t-1}, s_t = 0)), \boldsymbol{\sigma}) \tag{A.22}$$

Consider next the program

$$\begin{aligned} \mathcal{L}^*(x) = \quad &\max_{y,\boldsymbol{\sigma},\boldsymbol{\sigma'}} \quad V(x, y, \boldsymbol{\sigma}) \ - \ V(x, h(x), \boldsymbol{\sigma'}) \\ &\text{s.t.} \quad x \leq y \leq z(x) \end{aligned} \tag{A.23}$$

Since we are by assumption in Case 1 and hence t's posterior can be assumed to be arbitrarily close to 1, using (A.21) and (A.22) we can now see that to prove Proposition 2 it is enough to show that $\mathcal{L}^*(x)$ defined in (A.23) is negative for x sufficiently close to 1.

To solve (A.23), we form the Lagrangian

$$\mathcal{L}(x, y, \boldsymbol{\sigma}, \boldsymbol{\sigma'}, \phi, \psi) \;=\; [V(x, y, \boldsymbol{\sigma}) \;-\; V(x, h(x), \boldsymbol{\sigma'})] + \phi(z(x) - y) + \psi(y - x) \tag{A.24}$$

with Lagrange multipliers $\phi \geq 0$ and $\psi \geq 0$. Given x, a saddle point of $\mathcal{L}$ is defined as a list $(\bar{y}, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}', \bar{\phi}, \bar{\psi})$ that satisfies

$$\mathcal{L}(x, y, \boldsymbol{\sigma}, \boldsymbol{\sigma'}, \bar{\phi}, \bar{\psi}) \leq \mathcal{L}(x, \bar{y}, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}', \bar{\phi}, \bar{\psi}) \leq \mathcal{L}(x, \bar{y}, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}', \phi, \psi)$$

To see that saddle points exist, observe that $\mathcal{L}$ is continuous, the feasible sets of y, $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma'}$ are compact (in the latter two cases, in the product topology), and the constraints on y constitute a compact-valued correspondence of x. As always, $\mathcal{L}$ is linear in nonnegative multipliers so a minimizing pair $(\bar{\phi}, \bar{\psi})$ necessarily exists. Denote the set of saddle point solutions of $\mathcal{L}$ given x by $S^*(x) \times S_*(x)$ where $S^*(x)$ is the set of saddle maximizers $(\bar{y}, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}')$ and $S_*(x)$ the set of saddle minimizers $(\bar{\phi}, \bar{\psi})$. By Berge's Maximum Theorem (Berge, 1963), $S^*(x)$ and $S_*(x)$ are upper-hemicontinuous correspondences, and $S^*(x)$ is compact valued. Finally, it is well known (see for instance Rockafellar, 1970) that saddle points of $\mathcal{L}$ are solutions to the original problem in (A.23), and so for any saddle point, $(\bar{y}, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}', \bar{\phi}, \bar{\psi})$ we must have

$$\mathcal{L}^*(x) \;\equiv\; \mathcal{L}(x, \bar{y}, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}', \bar{\phi}, \bar{\psi})$$

Therefore, it now suffices to show that $\mathcal{L}(x, \bar{y}, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}', \bar{\phi}, \bar{\psi}) < 0$ when x is sufficiently close to one.

To see that this is the case, observe first that the derivative $\mathcal{L}_x$ of $\mathcal{L}$ with respect to x clearly exists, and $\mathcal{L}_x(x, y, \boldsymbol{\sigma}, \boldsymbol{\sigma'}, \phi, \psi)$ is a continuous function of all its arguments. Consequently, an Envelope Theorem due to Milgrom and Segal (2002, Theorem 5) can be applied to show that $\mathcal{L}^*(x)$ is left-differentiable with left-derivative given by

$$\mathcal{L}^*_{x^-}(x) = \min_{(\bar{y}, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}') \in S^*(x)} \; \max_{(\bar{\phi}, \bar{\psi}) \in S_*(x)} \; \mathcal{L}_x(x, \bar{y}, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}', \bar{\phi}, \bar{\psi}) \tag{A.25}$$

provided that the right-hand side optimization problem in (A.25) has a solution.

The value at $x = 1$ of the derivative $\mathcal{L}_x$ on the right-hand side of (A.25) after some tedious algebra can be seen to be equal to,

$$\mathcal{L}_x(1, 1, \overline{\boldsymbol{\sigma}}, \overline{\boldsymbol{\sigma}}', \bar{\phi}, \bar{\psi}) = -2\beta h'(1) + \bar{\phi} z'(1) - \bar{\psi}$$

Observe that this derivative does not depend on $\overline{\boldsymbol{\sigma}}$ and $\overline{\boldsymbol{\sigma}}'$ and that $\bar{y} = 1$ follows from the constraint in problem (A.23) and (A.20). Hence, when $x = 1$, the left-derivative in (A.25) can now be seen to be equal to

$$\mathcal{L}^*_{x^-}(1) = \max_{(\bar{\phi}, \bar{\psi}) \in S_*(1)} [-2\beta h'(1) + \bar{\phi} z'(1) - \bar{\psi}] \tag{A.26}$$

We need to show that (A.26) has a well defined solution. From the first order necessary conditions for the saddle point of problem (A.23) (in other words equating to zero the partial derivative of $\mathcal{L}$ with respect to y when $x = 1$) we obtain

$$\bar{\phi} - \bar{\psi} = 2\beta \tag{A.27}$$

Substituting (A.27) into (A.26) gives

$$\max_{S_*(1)} \; 2\beta(z'(1) - h'(1)) + \bar{\psi}(z'(1) - 1)$$

Notice that, using (A.20), $z'(1) - h'(1) > 0$ while $z'(1) - 1 < 0$.

We now proceed to prove that $\bar{\psi} = 0$ must be a solution in $S_*(1)$. To see that this is the case, recall that $S_*(\cdot)$ is upper hemicontinuous by Berge's Maximum Theorem (Berge, 1963), and so it suffices to show that $\bar{\psi} = 0$ is a solution in $S_*(1-\varepsilon)$ for all $\varepsilon > 0$ sufficiently small. Consider then any $x < 1$. In this case it cannot be that both constraints bind. That is, since $z(x) > x$ for all $x < 1$, the constraint $x \leq y \leq z(x)$ means that either $\bar{\phi} = 0$ or $\bar{\psi} = 0$ (or both) whenever $x < 1$. But for $x = 1 - \varepsilon$, the first order condition for y implies that for some $f(\varepsilon)$ for which $\lim_{\varepsilon \to 0} f(\varepsilon) = 0$

$$\bar{\phi} - \bar{\psi} > 2\beta + f(\varepsilon)$$

which for ε small, cannot be satisfied when $\bar{\phi} = 0$ and $\bar{\psi} > 0$. It is therefore only satisfied when $\bar{\psi} = 0$. Hence, $\bar{\psi} = 0 \in S_*(1)$ and so,

$$\max_{S_*(1)} \; 2\beta(z'(1) - h'(1)) + \bar{\psi}(z'(1) - 1) \; = \; 2\beta(z'(1) - h'(1)) \; > \; 0$$

Therefore $\mathcal{L}^*_{x^-}(1) > 0$, as required. Since $\mathcal{L}^*(1) = 0$, we conclude that $\mathcal{L}^*(x) < 0$ for x sufficiently close to one, and this concludes the proof of Proposition 2. ■

A.3. Proof of Proposition 3

The proof is constructive and it starts with a few preliminaries that will be used in the construction of the message spaces and equilibria that induce limit learning.

Fix a prior r and a signal quality p. Suppose that someone could observe a signal profile of arbitrary length, comprising k_1 realizations of value 1 and k_0 realizations of value 0.

Let $k = k_1 - k_0$. Then, using Bayes' rule and simple algebra, the belief of this observer that $\omega = 1$ would be given by

$$\gamma(r, p, k) \; = \; \frac{1}{1 + \dfrac{1-r}{r}\left(\dfrac{1-p}{p}\right)^k}$$

Hence,

$$\Gamma(r, p) \; = \; \{x \in (0, 1) \mid x \; = \; \gamma(r, p, k) \quad \text{for some} \quad k = 0, \pm 1, \pm 2, \dots\} \tag{A.28}$$

denotes the set of beliefs that can be generated by observing arbitrary signal profiles of any length given r and p.

Let z denote the largest element of $\Gamma(r, p)$ not exceeding $1/2$. In other words, let $z = \max\{x \in \Gamma(r, p) \mid x \in (0, 1/2]\}$. Consider a player with a beginning-of-period belief equal to z. After observing signals zero and one his beliefs are denoted z_0 and z_1 respectively. Using Bayes' rule, these can be trivially computed as

$$z_0 = \frac{z(1-p)}{z(1-p) + (1-z)p} \qquad \text{and} \qquad z_1 = \frac{zp}{zp + (1-z)(1-p)} \tag{A.29}$$

From (A.29) and the way we have defined z it is evident that $0 < z_0 \leq 1 - p$ and $1/2 < z_1 \leq p$. Since by assumption $\beta < p - 1/2$ it then follows easily that for some $\eta > 0$ it must be that for every $a' \in [-\beta, \eta - \beta]$

and every $a'' \in [1-\beta-\eta, 1-\beta]$ we have that both of the following inequalities hold

$$\begin{aligned} -z_0(1-a')^2 - (1-z_0)a'^2 &> -z_0(1-a'')^2 - (1-z_0)a''^2 \\ -z_1(1-a')^2 - (1-z_1)a'^2 &< -z_1(1-a'')^2 - (1-z_1)a''^2 \end{aligned} \tag{A.30}$$

In other words, any player who believes that $\omega = 1$ with probability z_0 prefers any $a' \in [-\beta, \eta - \beta]$ to any $a'' \in [1-\beta-\eta, 1-\beta]$. Moreover these preferences are reversed for any player who believes that $\omega = 1$ with probability z_1.

It useful to observe at this point that since $z \in \Gamma(r,p)$, there must exist an integer K (positive or negative) that satisfies

$$z = \frac{rp^K}{rp^K + (1-r)\,(1-p)^K} \tag{A.31}$$

For the rest of the proof we let T be an integer such that $T - 1 + K$ is an even positive integer.

Our construction relies on strategies that embody a "review phase" of length T, with T large. During this review phase the strategies keep track of the difference between the number of signals equal to one minus the number of signals equal to zero. If this difference is more than K, then the "outcome" of the review phase is "one," while if the difference is less than or equal to K then the "outcome" of the review is "zero."

It turns out that it is convenient for the argument that follows to write the basics of the review phase in terms of T and the number of signals equal to one, rather than in terms of the difference we have just mentioned. This is why we define the review phase "outcome function" $R_T : \{0,1\}^T \to \{0,1\}$ as follows. Consider an arbitrary sequence of signals $s^T = (s_1, \ldots, s_T)$ of length T, and let

$$R_T(s^T) = \begin{cases} 1 & \text{if} \quad \displaystyle\sum_{t=1}^{T} s_t > \frac{T-1+K}{2} \\ 0 & \text{if} \quad \displaystyle\sum_{t=1}^{T} s_t \leq \frac{T-1+K}{2} \end{cases}$$

The Weak Law of Large Numbers implies that for every $\varepsilon > 0$, there exists a $\bar{T}$ such that for every $T > \bar{T}$

$$\Pr[R_T(s^T) \,=\, i \mid \omega = i] \;>\; 1 - \varepsilon \qquad \forall\, i \in \{0,1\} \tag{A.32}$$

and

$$\Pr[\omega \,=\, i \mid R_T(s^T) \,=\, i] \;>\; 1 - \varepsilon \qquad \forall\, i \in \{0,1\} \tag{A.33}$$

We are now ready to specify the message spaces and strategies that induce limit learning of the state as required. Note that these will not depend on δ even though this is allowed in principle by the statement of Proposition 3 (see footnote 18).

Fix an arbitrary $\varepsilon > 0$ and an $\eta > 0$ as in (A.30). Then using (A.33) find an integer T that satisfies (A.32) and such that

$$\bar{x}_1 \;=\; \Pr[\omega = 1 \,|\, R_T(s^T) \;=\; 1] > \max\{1-\varepsilon, 1-\eta\}$$

$$\bar{x}_0 \;=\; \Pr[\omega = 1 \,|\, R_T(s^T) \;=\; 0] < \min\{\varepsilon, \eta\}$$

and

$$\frac{\bar{x}_1(1-p)}{\bar{x}_1(1-p)+(1-\bar{x}_1)p} > z_1 \qquad\qquad \frac{\bar{x}_0 p}{\bar{x}_0 p+(1-\bar{x}_0)(1-p)} < z_0 \tag{A.34}$$

Recall that our claim is that limit learning can obtain in a Sequential Equilibrium in the sense of Kreps and Wilson (1982) (see Subsection 3.1), rather than just in a Perfect Bayesian Equilibrium. We accomplish this by constructing the message spaces in a particularly careful way, so that once the strategies are specified *all* messages will be sent in equilibrium with positive probability. In this way there will be nothing further to prove since off-path beliefs will not even need to be defined. With this in mind, we are now ready to proceed with the construction of message spaces.

In addition to the standard messages corresponding to integers, consider two additional ones labeled 0^* and 1^*. The integer messages are interpreted as counting the number of signals equal to one, while the 0^* and 1^* are used to "declare" the outcome of the review phase. The five possibilities for M_t so that no messages are ever off-path are then as follows

(i) If $t < T,\ t \le \dfrac{T-1+K}{2}$ and $t < \dfrac{T+1-K}{2}$ then $M_t = \{0,1,\ldots,t\}$

(ii) If $t < T,\ t > \dfrac{T-1+K}{2}$ and $t < \dfrac{T+1-K}{2}$ then $M_t = \{0,1,\ldots,\dfrac{T-1+K}{2},1^*\}$

(iii) If $t < T,\ t \le \dfrac{T-1+K}{2}$ and $t \ge \dfrac{T+1-K}{2}$ then $M_t = \{t-\dfrac{T-1-K}{2},\ldots,t,0^*\}$

(iv) If $t < T,\ t > \dfrac{T-1+K}{2}$ and $t \ge \dfrac{T+1-K}{2}$ then $M_t = \{t-\dfrac{T-1-K}{2},\ldots,\dfrac{T-1+K}{2},0^*,1^*\}$

(v) If $t \ge T$ then $M_t = \{0^*,1^*\}$

We are now ready to construct a (pure) equilibrium strategy profile $\boldsymbol{\sigma}^*$ that induces limit learning of the state as required. Intuitively, all players $t \le T$ truthfully participate in the review phase. They truthfully count the tally of signals equal to one they are given by their predecessors on the basis of the actual signal they observe. (Player 1 has no predecessor, so he just counts to one or to zero according to the signal he observes.) Once the review phase is declared either by 0^* or 1^* all players babble, indefinitely into the future.

Formally, $\sigma_1^*(m_0,s_1) = s_1$ for all $s_1 \in \{0,1\}$. For all $t \ge 2$, the strategy σ_t^* is then defined by

$$\sigma_t^*(m_{t-1},s_t) = \begin{cases} m_{t-1} & \text{if } m_{t-1} \in \{0^*,1^*\}, \text{ regardless of } s_t \\ m_{t-1}+1 & \text{if } m_{t-1}+1 \leqslant \dfrac{T-1+K}{2} \text{ and } s_t = 1 \\ 1^* & \text{if } m_{t-1} = \dfrac{T-1+K}{2} \text{ and } s_t = 1 \\ m_{t-1} & \text{if } m_{t-1} > t + \dfrac{K-T-1}{2} \text{ and } s_t = 0 \\ 0^* & \text{if } m_{t-1} \le t + \dfrac{K-T-1}{2} \text{ and } s_t = 0 \end{cases} \tag{A.35}$$

Observe that if the players use the strategy profile $\boldsymbol{\sigma}^*$ defined in (A.35), then, using (A.32) and (A.33), in any period $t > T$ the beginning-of-period belief x_t will be greater than $1-\varepsilon$ (smaller than ε) with probability at least $1-\varepsilon$ when the state is $\omega = 1$ ($\omega = 0$). Hence limit learning obtains under $\boldsymbol{\sigma}^*$, as required.

It remains to check that $\boldsymbol{\sigma}^*$ is in fact an equilibrium for δ sufficiently large.

Notice that in any period $t \ge T$ only two messages will ever be sent, and only the two actions $\bar{x}_0 - \beta$

and $\bar{x}_1 - \beta$ will ever be chosen. It then follows easily from (A.30) and (A.34) that no player wants to deviate from the proposed equilibrium after T.

Finally, consider any player $t \leq T$. If he follows the proposed equilibrium strategy defined in (A.35), from $T+1$ onwards the following will take place. The action $\bar{x}_1 - \beta$ will be played if the prior r updated on the basis of the first T signal realizations yields a belief at least as large as z_1. Conversely, $\bar{x}_0 - \beta$ will be played if the prior r updated on the basis of the first T signal realizations yields a belief at most equal to z_0. These two actions are the only ones that will possibly be taken in period $T+1$ and all subsequent periods. Hence it follows from (A.30) that a player t who plays according to σ_t^* will induce his preferred action among these two with probability one in every period greater than T. Deviating from σ_t^* will necessarily reduce this probability to be strictly less than one. While following the equilibrium strategy might be costly for player t in periods before T, this is clearly sufficient to show that he will not want to deviate from the proposed equilibrium if δ is close enough to one. ■

A.4. Proof of Proposition 4

The proof proceeds by contradiction. If limit learning obtains, some player must have a sufficiently extreme beginning-of-period belief. We show that this player has a strict incentive to announce that $\omega = 1$ after both realizations of the signal. This contradicts limit learning and hence allows us to close the argument.

As in the proof of Proposition 2, without loss of generality, throughout the argument we consider strategies such that for every m^{t-1}, $x_{t+1}\left(m^{t-1}, m_t = 1\right) \geqslant x_{t+1}\left(m^{t-1}, m_t = 0\right)$.

Recall that to prove Proposition 2, we showed that any equilibrium strategy profile $\boldsymbol{\sigma}$ must have beliefs that are bounded above according to Definition A.2. As we also remarked in the proof of Proposition 2, a similar argument can be used to establish that any equilibrium strategy profile $\boldsymbol{\sigma}$ must have beliefs that are bounded below according to Definition A.2.

In the remainder of the proof of Proposition 4 we use the fact that any equilibrium strategy profile $\boldsymbol{\sigma}$ must have beliefs that are bounded below. In other words, from now on we will take it as given that given any equilibrium $\boldsymbol{\sigma}$ we can find a greatest lower bound $\underline{x} > 0$ on the beginning-of-period beliefs.

Using the fact that by assumption $p < \sqrt{2}/(1+\sqrt{2})$, simple algebra is then sufficient to show that there exists an $\eta > 0$ such that for any $\hat{x} \in (0, \eta)$, any $\tilde{x} > \hat{x}$ and any x satisfying

$$\frac{\tilde{x}(1-p)^2}{\tilde{x}(1-p)^2 + (1-\tilde{x})p^2} \leq x \leq \frac{\tilde{x}(1-p)}{\tilde{x}(1-p) + (1-\tilde{x})p}$$

we have that

$$-x[1-\tilde{x}+\beta]^2 - (1-x)[\tilde{x}-\beta]^2 > -x[1-\hat{x}+\beta]^2 - (1-x)[\hat{x}-\beta]^2 \tag{A.36}$$

Now suppose by way of contradiction that limit learning of the state is possible under the hypotheses of the proposition. Let $\boldsymbol{\sigma}^\delta$ be the sequence of equilibria in which limit learning takes place. For each such equilibrium, let $\underline{x}_\delta$ be the greatest lower bound on the beginning-of-period that, as we asserted above, we know can be found.

Since η as needed for (A.36) to hold is in fact fixed, it follows from our contradiction hypothesis that choosing δ sufficiently large we can ensure that $\underline{x}_\delta < \eta$.

In the rest of the proof we concentrate on the case in which the greatest lower bound $\underline{x}_\delta$ is actually achieved.[27] Considering the first (stochastic) time in which the greatest lower bound is achieved, we can find

[27] The argument for the case in which the beliefs converge to $\underline{x}_\delta$ without ever reaching this value is very close to the one we make explicit here. Some extra limit arguments are needed, which also involve extra notation. While the details are available from the authors upon request, we omit

a t and m^{t-1} such that $\underline{x}_\delta = x_{t+1}(m^{t-1}, m_t = 0)$, and, after observing m^{t-1}, player t's strategy satisfies

$$\sigma_t^\delta(m^{t-1}, s_t = 0) \neq \sigma_t^\delta(m^{t-1}, s_t = 1), \quad \sigma_t^\delta(m_t = 0|m^{t-1}, s_t = 0) > 0, \quad \sigma_t^\delta(m_t = 1|m^{t-1}, s_t = 1) > 0$$

which, in turn, imply that

$$x_t^E(m^{t-1}, s_t = 0) \leq \underline{x}_\delta = x_{t+1}(m^{t-1}, m_t = 0) <$$

$$x_t(m^{t-1}) = \frac{x_t^E(m^{t-1}, s_t = 0)p}{x_t^E(m^{t-1}, s_t = 0)p + (1 - x_t^E(m^{t-1}, s_t = 0))(1-p)} <$$

$$x_{t+1}(m^{t-1}, m_t = 1) \leq x_t^E(m^{t-1}, s_t = 1) = \frac{x_t^E(m^{t-1}, s_t = 0)p^2}{x_t^E(m^{t-1}, s_t = 0)p^2 + (1 - x_t^E(m^{t-1}, s_t = 0))(1-p)^2}$$

Now consider the incentives of player t after observing $s_t = 0$. If he announces the message $m_t = 0$ (recall that in equilibrium this happens with positive probability), then the action $\underline{x}_\delta - \beta$ will be chosen in every subsequent period. This is because the lower bound on beliefs $\underline{x}_\delta$ will have been achieved, and this implies that all players will babble in the continuation equilibrium prescribed by $\boldsymbol{\sigma}^\delta$.

Suppose now instead that player t after observing $s_t = 0$, and therefore assigning probability $x_t^E(m^{t-1}, s_t = 0)$ to the state $\omega = 1$, reports the message $m_t = 1$ (again, notice that in equilibrium $m_t = 1$ is sent with positive probability by player t after observing $s_t = 1$).

Consider now an arbitrary period $t + \tau$ with $\tau \geqslant 1$ and an arbitrary sequence of messages $m^{t,t+\tau-1}$ with $m_t = 0$ which occurs with positive probability after the initial sequence of messages m^{t-1}. Two cases are possible. The first case is that in period $t + \tau$ the action $\underline{x}_\delta - \beta$ is chosen because the belief has reached its lower bound $\underline{x}_\delta$. In this case player t is clearly indifferent between reporting $m_t = 0$ or $m_t = 1$.

The second case is that the sequence of messages $m^{t,t+\tau-1}$ is such that by period $t+\tau$ the lower bound for beliefs has not been reached and perpetual babbling has not necessarily begun: $x_{t+\tau}\left(m^{t-1}, m^{t,t+\tau-1}\right) > \underline{x}_\delta$. How does player t who observes $s_t = 0$ evaluate this situation? At this point tis update belief x must satisfy

$$\frac{x_{t+\tau}(m^{t-1}, m^{t,t+\tau-1})(1-p)^2}{x_{t+\tau}(m^{t-1}, m^{t,t+\tau-1})(1-p)^2 + (1 - x_{t+\tau}(m^{t-1}, m^{t,t+\tau-1}))p^2} \leq x \leq$$

$$\frac{x_{t+\tau}(m^{t-1}, m^{t,t+\tau-1})(1-p)}{x_{t+\tau}(m^{t-1}, m^{t,t+\tau-1})(1-p) + (1 - x_{t+\tau}(m^{t-1}, m^{t,t+\tau-1}))p}$$

Given this belief, it follows from (A.36) that player t strictly prefers action $x_{t+\tau}\left(m^{t-1}, m^{t,t+\tau-1}\right) - \beta$ to action $\underline{x}_\delta - \beta$.

We conclude that for player t who observes $s_t = 0$ sending $m_t = 1$ is strictly better than sending $m_t = 0$ in every period $t+1, t+2, \ldots$ provided that the lower bound on beliefs has not been reached. Since the probability that the bound has not been reached in any period is obviously strictly positive, we can then conclude that the player has a strict incentive to send $m_t = 1$, deviating from what the putative equilibrium $\boldsymbol{\sigma}^\delta$ prescribes. Hence, the proof of Proposition 4 is now complete. ■

this case for the sake of brevity.

References

Anderlini, L., D. Gerardi, and R. Lagunoff (2008): "A 'Super' Folk Theorem for Dynastic Repeated Games," *Economic Theory*, 37, 357–394.

——— (2010): "Social Memory, Evidence and Conflict," *Review of Economic Dynamics*, 13, 559–574.

Anderlini, L., and R. Lagunoff (2005): "Communication in Dynastic Repeated Games: 'Whitewashes' and 'Coverups'," *Economic Theory*, 26, 265–299.

Banerjee, A. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817.

Berge, C. (1963): *Topological Spaces*. New York: Macmillan.

Carrillo, J., and T. Mariotti (2000): "Strategic Ignorance as a Self-Disciplining Device," *Review of Economic Studies*, 67, 529–44.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1351.

Farrell, J., and M. Rabin (1996): "Cheap Talk," *Journal of Economic Perspectives*, 10, 103–118.

Forges, F. (1986): "An Approach to Communication Equilibria," *Econometrica*, 54, 1375–1385.

Fudenberg, D., and J. Tirole (1991): *Game Theory*. Cambridge, Massachusetts: The MIT Press.

Gale, D. (1996): "What Have We Learned from Social Learning?," *European Economic Review, Papers and Proceedings of the Tenth Annual Congress of the European Economic Association*, 40, 617–628.

Golosov, M., V. Skreta, A. Tsyvinski, and A. Wilson (2011): "Dynamic Strategic Information Transmission," NYU, mimeo.

Hawking, S. W. (ed.) (2003): *On the Shoulders of Giants: The Great Works of Physics and Astronomy*. Philadelphia: Running Press.

Kartik, N. (2009): "Strategic Communication with Lying Costs," *Review of Economics Studies*, 76, 1359–1395.

Kobayashi, H. (2007): "Folk Theorems for Infinitely Repeated Games Played by Organizations with Short-Lived Members," *International Economic Review*, 48, 517–549.

Kreps, D. M., and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

Krishna, V., and J. Morgan (2001): "A Model of Expertise," *Quarterly Journal of Economics*, 116, 747–775.

——— (2008): "Cheap Talk," in *The New Palgrave Dictionary of Economics*, ed. by L. Blume, and S. Durlauf, 2^{nd} edn. New York, NY: McMillan.

Lagunoff, R. (2006): "Credible Communication in Dynastic Government," *Journal of Public Economics*, 90, 59–86.

Lagunoff, R., and A. Matsui (2004): "Organizations and Overlapping Generations Games: Memory, Communication, and Altruism," *Review of Economic Design*, 8, 383–411.

Li, W. (2007): "Changing One's Mind when the Facts Change: Incentives of Experts and the Design of Reporting Protocols," *Review of Economic Studies*, 74, 1175–1194.

Mailath, G., and L. Samuelson (2006): *Repeated Games and Reputations: Long Run Relationships*. Oxford: Oxford University Press.

Milgrom, P., and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

Moscarini, G., M. Ottaviani, and L. Smith (1998): "Social Learning in a Changing World," *Economic Theory*, 11, 657–665.

Myerson, R. B. (1982): "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10, 67–81.

Ottaviani, M., and P. N. Sørensen (2006): "Professional Advice," *Journal of Economic Theory*, 126, 120–142.

Rockafellar, R. T. (1970): *Convex Analysis*. Princeton, NJ: Princeton University Press.

Smith, L., and P. Sørensen (2008): "Informational Herding and Optimal Experimentation," University of Michigan, mimeo.

Sobel, J. (2000): "Economists' Models of Learning," *Journal of Economic Theory*, 94, 241–261.

——— (2009): "Signaling Games," in *Encyclopedia of Complexity and System Science*, ed. by R. A. Myers. Berlin: Springer-Verlag.

Spector, D. (2000): "Rational Debate and One-Dimensional Conflict," *Quarterly Journal of Economics*, 115, 181–200.

Taub, B. (1997): "Dynamic Agency with Feedback," *Rand Journal of Economics*, 28, 515–543.

Supplement to "Communication and Learning:" Omitted Proofs

Luca Anderlini   Dino Gerardi   Roger Lagunoff

June 2011

Proof of Lemma A.2: Let $\hat{U}_t(m_t, \omega)$ be the same type of continuation payoff as $U_t(m_{t-1}, \omega)$ defined in (A.2), but this time assuming that t *sends* message m_t. Clearly, $\hat{U}_t(m_t, \omega)$ and $U_t(m_{t-1}, \omega)$ are related so that

$$U_t(m_{t-1}, \omega) = \int\limits_{\gamma_t \in [0,1]} [\Pr(s_t = 0|\omega)\hat{U}_t(\sigma_t(m_{t-1}, s_t = 0, \gamma_t), \omega) + \Pr(s_t = 1|\omega)\hat{U}_t(\sigma_t(m_{t-1}, s_t = 1, \gamma_t), \omega)] \, \mathrm{d}\, \gamma_t$$

Fix two messages m_{t-1} and m'_{t-1} with $x_t(m_{t-1}) > x_t(m'_{t-1})$. Consider any $\tau > 0$ and any sequence $s^{t,t+\tau-1}$. We define the set

$$Z^{t,t+\tau-1}(m_{t-1}, m'_{t-1}, s^{t,t+\tau-1}) = \Big\{ \gamma^{t,t+\tau-1} \in [0,1]^\tau \text{ such that } \mathrm{a}_{t+\tau'}(m_{t-1}, s^{t,t+\tau'-1}, \gamma^{t,t+\tau'-1}) > \mathrm{a}_{t+\tau'}(m'_{t-1}, s^{t,t+\tau'-1}, \gamma^{t,t+\tau'-1}) \ \ \forall\, \tau' = 1, ..., \tau \Big\}$$

For convenience, the rest of the argument is divided into three separate steps.

Step 1: *Fix an M and σ^* as in Remark A.1. Let $m_{\bar{t}-1}$ and $m'_{\bar{t}-1}$ be two messages such that $x_{\bar{t}}(m_{\bar{t}-1}) > x_{\bar{t}}\left(m'_{\bar{t}-1}\right)$. Then for every $\omega \in \{0, 1\}$ it must be that*

$$\begin{aligned} & U_{\bar{t}}(m_{\bar{t}-1}, \omega) - U_{\bar{t}}(m'_{\bar{t}-1}, \omega) = \\ & \quad (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} \sum_{s^{\bar{t},\bar{t}+t-1} \in \{0,1\}^t} \Pr(s^{\bar{t},\bar{t}+t-1} \,|\, \omega) \\ & \quad \int\limits_{Z^{\bar{t},\bar{t}+t-1}(m_{t-1}, m'_{t-1}, s^{\bar{t},\bar{t}+t-1})} [(\omega - \mathrm{a}_{\bar{t}+t}(m_{t-1}, s^{\bar{t},\bar{t}+t-1}, \gamma^{\bar{t},\bar{t}+t-1}))^2 - (\omega - \mathrm{a}_{\bar{t}+t}(m_{t-1},' s^{\bar{t},\bar{t}+t-1}, \gamma^{\bar{t},\bar{t}+t-1}))^2] \, \mathrm{d}\, \gamma^{\bar{t},\bar{t}+t-1} \end{aligned} \tag{S.1}$$

To keep notation usage down, during the proof of this step we let $m_{\bar{t}-1} = m$ and $m'_{\bar{t}-1} = m'$.

Suppose that for some $t \geq 0$ there is a pair of sequences $s^{\bar{t},\bar{t}+t} = (s_{\bar{t}}, \ldots, s_{\bar{t}+t})$ and $\gamma^{\bar{t},\bar{t}+t} = (\gamma_{\bar{t}}, \ldots, \gamma_{\bar{t}+t})$ such that: (a) $\mathrm{a}_{\bar{t}+\tau}(m, s_{\bar{t}}, \ldots, s_{\bar{t}+\tau-1}, \gamma_{\bar{t}}, \ldots, \gamma_{\bar{t}+\tau-1}) > \mathrm{a}_{\bar{t}+\tau}(m', s_{\bar{t}}, \ldots, s_{\bar{t}+\tau-1}, \gamma_{\bar{t}}, \ldots, \gamma_{\bar{t}+\tau-1})$ for every $\tau \leq t$, and (b) $\mathrm{a}_{\bar{t}+t+1}(m, s_{\bar{t}}, \ldots, s_{\bar{t}+t}, \gamma_{\bar{t}}, \ldots, \gamma_{\bar{t}+t}) \leq \mathrm{a}_{\bar{t}+t+1}(m', s_{\bar{t}}, \ldots, s_{\bar{t}+t}, \gamma_{\bar{t}}, \ldots, \gamma_{\bar{t}+t})$. (If such a pair of sequences cannot be found then there is nothing to prove.) We need to consider two, mutually exclusive and exhaustive, cases.

Case 1: After the sequences $s^{\bar{t},\bar{t}+t}$ and $\gamma^{\bar{t},\bar{t}+t}$, player $\bar{t}+t$ sends the same message (say $\tilde{m}$) both when player $\bar{t}$ behaves according to m and when player $\bar{t}$ behaves according to m'.

Clearly, after these sequences, the messages m and m' will induce the same action in every period $\bar{t}+t+\tau$ for any $\tau \geq 1$. Hence in Case 1 there is nothing further to prove.

Case 2: After the sequences $s^{\bar{t},\bar{t}+t}$ and $\gamma^{\bar{t},\bar{t}+t}$, player $\bar{t}+t$ sends message $\tilde{m}$ if player $\bar{t}$ behaves according to m and message $\tilde{m}' \neq \tilde{m}$ if player $\bar{t}$ behaves according to m'.

Let $\tilde{y}$ (resp. $\tilde{y}'$) denote the beginning-of-period belief of player $\bar{t}+t+1$ when he receives $\tilde{m}$ (resp. $\tilde{m}'$). Of course, $\tilde{y}' \geq \tilde{y}$ since

$$\tilde{y} - \beta \; = \; \mathrm{a}_{\bar{t}+t+1}(m, s^{\bar{t},\bar{t}+t}, \gamma^{\bar{t},\bar{t}+t}) \; \leq \; \mathrm{a}_{\bar{t}+t+1}(m', s^{\bar{t},\bar{t}+t}, \gamma^{\bar{t},\bar{t}+t}) \; = \; \tilde{y}' - \beta$$

Let $\tilde{x}$ denote the end-of-period belief of player $\bar{t}+t$ when player $\bar{t}$ behaves according to m and the sequences of realized signals and randomization devices are $s^{\bar{t},\bar{t}+t}$ and $\gamma^{\bar{t},\bar{t}+t}$, respectively. Finally, let $\tilde{x}'$ denote the end-of-period belief of player $\bar{t}+t$ when player $\bar{t}$ behaves according to m' and the sequences are $s^{\bar{t},\bar{t}+t}$ and $\gamma^{\bar{t},\bar{t}+t}$. Notice that $\tilde{x} > \tilde{x}'$ since by assumption $\mathrm{a}_{\bar{t}+t}(m, s_{\bar{t}+1}, \ldots, s_{\bar{t}+t-1}, \gamma_{\bar{t}+1}, \ldots, \gamma_{\bar{t}+t-1}) > \mathrm{a}_{\bar{t}+t}(m', s_{\bar{t}+1}, \ldots, s_{\bar{t}+t-1}, \gamma_{\bar{t}+1}, \ldots, \gamma_{\bar{t}+t-1})$.

Let $\mathcal{T}_{\bar{t}+t} \subset M_{\bar{t}+t-1} \times \{0,1\} \times [0,1]$ denote the set of types (a type consisting of the message received and the two random variables observed) of player $\bar{t}+t$ who send message $\tilde{m}$. Also, let $\tilde{X}^E_{\bar{t}+t}$ denote the set of corresponding beliefs, so that $\tilde{X}^E_{\bar{t}+t} = \bigcup_{(m_{\bar{t}+t-1}, s_{\bar{t}+t}, \gamma_{\bar{t}+t}) \in \mathcal{T}_{\bar{t}+t}} x^E_{\bar{t}+t}(m_{\bar{t}+t-1}, s_{\bar{t}+t}, \gamma_{\bar{t}+t})$. Define $\mathcal{T}'_{\bar{t}+t}$ and $\tilde{X}^{E'}_{\bar{t}+t}$ in a similar way (replace $\tilde{m}$ with $\tilde{m}'$).

Since each player uses Bayes' rule to compute his beginning-of-period beliefs, $\tilde{y}$ belongs to the convex hull of $\tilde{X}^E_{\bar{t}+t}$ and $\tilde{y}'$ belongs to the convex hull of $\tilde{X}^{E'}_{\bar{t}+t}$. Notice that $\tilde{x} \in \tilde{X}^E_{\bar{t}+t}$ and $\tilde{x}' \in \tilde{X}^{E'}_{\bar{t}+t}$, and recall that $\tilde{x} > \tilde{x}'$. This, together with $\tilde{y}' \geq \tilde{y}$, imply that one of the following two mutually exclusive subcases, (a) and (b), must be true.

(a) We can find three types $(m^1_{\bar{t}+t-1}, s^1_{\bar{t}+t}, \gamma^1_{\bar{t}+t})$, $(m^2_{\bar{t}+t-1}, s^2_{\bar{t}+t}, \gamma^2_{\bar{t}+t})$ and $(m^3_{\bar{t}+t-1}, s^3_{\bar{t}+t}, \gamma^3_{\bar{t}+t})$ such that

$$x^{(1)} = x^E_{\bar{t}+t}(m^1_{\bar{t}+t-1}, s^1_{\bar{t}+t}, \gamma^1_{\bar{t}+t}) < x^{(2)} = x^E_{\bar{t}+t}(m^2_{\bar{t}+t-1}, s^2_{\bar{t}+t}, \gamma^2_{\bar{t}+t}) < x^{(3)} = x^E_{\bar{t}+t}(m^3_{\bar{t}+t-1}, s^3_{\bar{t}+t}, \gamma^3_{\bar{t}+t}) \quad \text{(S.2)}$$

and, the two extreme types $(m^1_{\bar{t}+t-1}, s^1_{\bar{t}+t}, \gamma^1_{\bar{t}+t})$ and $(m^3_{\bar{t}+t-1}, s^3_{\bar{t}+t}, \gamma^3_{\bar{t}+t})$ send the same message, equal to either $\tilde{m}$ or $\tilde{m}'$, while the intermediate type sends the other. Suppose that the extreme types send $\tilde{m}$ and $(m^2_{\bar{t}+t-1}, s^2_{\bar{t}+t}, \gamma^2_{\bar{t}+t})$ sends $\tilde{m}'$ (mutatis mutandis, the reverse case is identical). Since in equilibrium no type can have a profitable deviation when selecting which message to send, it must be that for $k = 1, 3$

$$x^{(k)} \hat{U}_{\bar{t}+t}(\tilde{m}, 1) + (1 - x^{(k)}) \hat{U}_{\bar{t}+t}(\tilde{m}, 0) \geq x^{(k)} \hat{U}_{\bar{t}+t}(\tilde{m}', 1) + (1 - x^{(k)}) \hat{U}_{\bar{t}+t}(\tilde{m}', 0) \quad \text{(S.3)}$$

and

$$x^{(2)} \hat{U}_{\bar{t}+t}(\tilde{m}, 1) + (1 - x^{(2)}) \hat{U}_{\bar{t}+t}(\tilde{m}, 0) \leq x^{(2)} \hat{U}_{\bar{t}+t}(\tilde{m}', 1) + (1 - x^{(2)}) \hat{U}_{\bar{t}+t}(\tilde{m}', 0) \quad \text{(S.4)}$$

Recall that by (S.2) we have $x^{(1)} < x^{(2)} < x^{(3)}$. Thus inequalities (S.3) and (S.4) can only be satisfied if $\hat{U}_{\bar{t}+t}(\tilde{m}, \omega) = \hat{U}_{\bar{t}+t}(\tilde{m}', \omega)$, $\forall\, \omega \in \{0,1\}$. Therefore, after the sequences $s^{\bar{t},\bar{t}+t}$ and $\gamma^{\bar{t},\bar{t}+t}$, player $\bar{t}$ receives the same expected continuation payoff regardless of ω and regardless of whether he behaves according to m or behaves according to m'. This concludes the argument in subcase (a).

(b) There are four types, $(m^1_{\bar{t}+t-1}, s^1_{\bar{t}+t}, \gamma^1_{\bar{t}+t})$, $(m^2_{\bar{t}+t-1}, s^2_{\bar{t}+t}, \gamma^2_{\bar{t}+t})$, $(m^3_{\bar{t}+t-1}, s^3_{\bar{t}+t}, \gamma^3_{\bar{t}+t})$ and $(m^4_{\bar{t}+t-1}, s^4_{\bar{t}+t}, \gamma^4_{\bar{t}+t})$ such that

$$x^E_{\bar{t}+t}(m^1_{\bar{t}+t-1}, s^1_{\bar{t}+t}, \gamma^1_{\bar{t}+t}) = x^E_{\bar{t}+t}(m^2_{\bar{t}+t-1}, s^2_{\bar{t}+t}, \gamma^2_{\bar{t}+t}) = \tilde{x}$$

and

$$x^E_{\bar{t}+t}(m^3_{\bar{t}+t-1}, s^3_{\bar{t}+t}, \gamma^3_{\bar{t}+t}) = x^E_{\bar{t}+t}(m^4_{\bar{t}+t-1}, s^4_{\bar{t}+t}, \gamma^3_{\bar{t}+t}) = \tilde{x}'$$

and the two types $(m^1_{\bar{t}+t-1}, s^1_{\bar{t}+t}, \gamma^1_{\bar{t}+t})$ and $(m^3_{\bar{t}+t-1}, s^3_{\bar{t}+t}, \gamma^3_{\bar{t}+t})$ send message $\tilde{m}$ while the two types $(m^2_{\bar{t}+t-1},$

$s^2_{\bar{t}+t}$, $\gamma^2_{\bar{t}+t}$) and $(m^4_{\bar{t}+t-1}, s^4_{\bar{t}+t}, \gamma^3_{\bar{t}+t})$ send message $\tilde{m}'$.

Again, in equilibrium no type can have a profitable deviation when selecting which message to send. Recalling that $\tilde{x} > \tilde{x}'$, it is then immediate to see that this implies that $\hat{U}_{\bar{t}+t}(\tilde{m},\omega) = \hat{U}_{\bar{t}+t}(\tilde{m}',\omega)$, $\forall\ \omega \in \{0,1\}$. Therefore, after the sequence $s^{\bar{t},\bar{t}+t}$ player $\bar{t}$ receives the same expected continuation payoff regardless of ω and regardless of whether he behaves according to m or behaves according to m'. This closes the argument in case (b) and hence concludes the proof of Step 1.

Step 2: *Fix an M and $\boldsymbol{\sigma}^*$ as in Remark A.1. For any $\eta > 0$ there exists an $\varepsilon > 0$ such that the following is true for every t. Suppose that m_{t-1} and m'_{t-1} are two messages in M_{t-1} with $\min\{x_t(m_{t-1}), x_t(m'_{t-1})\} > 1-\varepsilon$ and $x_t(m_{t-1}) \neq x_t(m'_{t-1})$. Then*

$$\left|U_t\left(m_{t-1},1\right) - U_t\left(m'_{t-1},1\right)\right| < \eta$$

Fix η and choose $\varepsilon > 0$ such that $\varepsilon(1-\beta)^2/(1-\varepsilon) < \eta$. By hypothesis, we can find two messages m_{t-1} and m'_{t-1} in M_{t-1} with $x_t(m_{t-1}) > 1-\varepsilon$ and $x_t(m'_{t-1}) > 1-\varepsilon$. To keep notation usage down, during the proof of this step we let $m_{t-1} = m$, $m'_{t-1} = m'$, $x_t(m_{t-1}) = x$ and $x_t(m'_{t-1}) = x'$. Without loss of generality, assume $x > x'$.

Since player t must have no incentives to deviate from his equilibrium strategy after observing s_t and γ_t, taking averages, the following two inequalities must be satisfied

$$x\, U_t(m,1) + (1-x)U_t(m,0) \geq x\, U_t(m',1) + (1-x)U_t(m',0)$$

and

$$x' U_t(m,1) + (1-x')U_t(m,0) \leq x' U_t(m',1) + (1-x')U_t(m',0)$$

Therefore, for some $\bar{x} \in [x', x]$ it must be that

$$\bar{x}\, U_t(m,1) + (1-\bar{x})U_t(m,0) = \bar{x}\, U_t(m',1) + (1-\bar{x})U_t(m',0)$$

Hence, using the fact that $\bar{x} > 1-\varepsilon$, we conclude that

$$|U_t(m,1) - U_t(m',1)| = \frac{(1-\bar{x})}{\bar{x}}|U_t(m',0) - U_t(m,0)| < \frac{\varepsilon}{1-\varepsilon}|U_t(m',0) - U_t(m,0)| \tag{S.5}$$

Notice now that, using (3), it is trivial that no player will ever choose an action outside $[-\beta, 1-\beta]$. From (A.2) it then follows directly that the continuation payoff $U_t(\cdot,0)$ is bounded above by 0 and below by $-(1-\beta)^2$. It is then obvious that $|U_t(m',0) - U_t(m,0)| < (1-\beta)^2$. Hence, because of the way ε was chosen, (S.5) is enough to prove the claim in Step 2.

Step 3: *Fix an M and $\boldsymbol{\sigma}^*$ as in Remark A.1. For any $\eta > 0$ there exists an $\varepsilon > 0$ such that the following is true for every t. Suppose that m_{t-1} and m'_{t-1} are two messages in M_{t-1} with $\min\{x_t(m_{t-1}), x_t(m'_{t-1})\} > 1-\varepsilon$ and $x_t(m_{t-1}) \neq x_t(m'_{t-1})$. Then*

$$|U_t(m_{t-1},0) - U_t(m'_{t-1},0)| < \eta$$

We proceed by contradiction. Then by hypothesis there must exist an η such that for every ε we can find messages m_{t-1} and m'_{t-1} (for some t) such that $\min\{x_t(m_{t-1}), x_t(m'_{t-1})\} > 1-\varepsilon$, $x_t(m_{t-1}) \neq x_t(m'_{t-1})$ and

$$|U_t(m_{t-1},0) - U_t(m'_{t-1},0)| > \eta \tag{S.6}$$

To keep notation usage down, during the proof of this step we let $m_{t-1} = m$, $m'_{t-1} = m'$, $x_t(m_{t-1}) = x$ and $x_t(m'_{t-1}) = x'$. Without loss of generality, assume $x > x'$. From Step 1 we know that

$$U_t(m,0) - U_t(m',0) =$$

$$(1-\delta)\sum_{\tau=1}^{\infty}\delta^{\tau-1}\sum_{s^{t,t+\tau-1}\in\{0,1\}^t}\Pr(s^{t,t+\tau-1}\,|\,\omega)$$
$$\int_{Z^{t,t+\tau-1}(m,m',s^{t,t+\tau-1})}\left[\mathrm{a}_{t+\tau}(m',s^{t,t+\tau-1},\gamma^{t,t+\tau-1})^2 - \mathrm{a}_{t+\tau}(m,s^{t,t+\tau-1},\gamma^{t,t+\tau-1})^2\right]\ \mathrm{d}\gamma^{t,t+\tau-1} \tag{S.7}$$

Just as in the proof of Step 2, the difference between continuation payoffs (appropriately normalized) conditional on $\omega = 0$ is bounded above by $(1-\beta)^2$. Therefore, from (S.7) we conclude that for any $T \geq 1$ it must be that

$$|U_t(m,0) - U_t(m',0)| \leq$$

$$(1-\delta)\sum_{\tau=1}^{T}\delta^{\tau-1}\sum_{s^{t,t+\tau-1}\in\{0,1t\}^t}\Pr(s^{t,t+\tau-1}|\omega)$$
$$\int_{Z^{t,t+\tau-1}(m,m',s^{t,t+\tau-1})}\left|\mathrm{a}_{t+\tau}(m',s^{t,t+\tau-1},\gamma^{t,t+\tau-1})^2 - \mathrm{a}_{t+\tau}(m,s^{t,t+\tau-1},\gamma^{t,t+\tau-1})^2\right|\ \mathrm{d}\,\gamma^{t,t+\tau-1}\ \ +\ \ \delta^T(1-\beta)^2$$

Since T can always be chosen so that $\delta^T(1-\beta)^2 < \eta/2$, we conclude that there exists a T such that

$$|U_t(m,0) - U_t(m',0)| <$$

$$(1-\delta)\sum_{\tau=1}^{T}\delta^{\tau-1}\sum_{s^{t,t+\tau-1}\in\{0,1\}^t}\Pr(s^{t,t+\tau-1}\,|\,\omega)$$
$$\int_{Z^{t,t+\tau-1}(m,m',s^{t,t+\tau-1})}\left|\mathrm{a}_{t+\tau}(m',s^{t,t+\tau-1},\gamma^{t,t+\tau-1})^2 - \mathrm{a}_{t+\tau}(m,s^{t,t+\tau-1},\gamma^{t,t+\tau-1})^2\right|\ \mathrm{d}\gamma^{t,t+\tau-1}\ +\ \frac{\eta}{2} \tag{S.8}$$

Inequalities (S.6) and (S.8) directly imply that

$$(1-\delta)\sum_{\tau=1}^{T}\delta^{\tau-1}\sum_{s^{t,t+\tau-1}\in\{0,1\}^t}\Pr(s^{t,t+\tau-1}\,|\,\omega)$$
$$\int_{Z^{t,t+\tau-1}(m,m',s^{t,t+\tau-1})}\left|\mathrm{a}_{t+\tau}(m',s^{t,t+\tau-1},\gamma^{t,t+\tau-1})^2 - \mathrm{a}_{t+\tau}(m,s^{t,t+\tau-1},\gamma^{t,t+\tau-1})^2\right|\ \mathrm{d}\,\gamma^{t,t+\tau-1}\ >\ \frac{\eta}{2} \tag{S.9}$$

However, inequality (S.9) implies that there exist a $\bar{\tau} = 1, \ldots, T$ and some sequence of signals $s^{t,t+\bar{\tau}-1}$

such that

$$\int_{Z^{t,t+\bar{\tau}-1}(m,m',s^{t,t+\bar{\tau}-1})} |\, \mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})^2 - \mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})^2 \,|\, \mathrm{d}\, \gamma^{t,t+\bar{\tau}-1} \; > \; \frac{\eta}{2(1-\delta^T)}$$

By definition, for any $\gamma^{t,t+\bar{\tau}-1} \in Z^{t,t+\bar{\tau}-1}(m, m', s^{t,t+\bar{\tau}-1})$, we have that $\mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1}) > \mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})$. Since all actions are weakly smaller than $1-\beta$, we have

$$\frac{\eta}{2(1-\delta^T)} < \int_{Z^{t,t+\bar{\tau}-1}(m,m',s^{t,t+\bar{\tau}-1})} |\mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})^2 - \mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})^2|\mathrm{d}\, \gamma^{t,t+\bar{\tau}-1} =$$

$$\int_{Z^{t,t+\bar{\tau}-1}(m,m',s^{t,t+\bar{\tau}-1})} |\mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1}) + \mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})|$$

$$[\mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1}) - \mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})]\, \mathrm{d}\gamma^{t,t+\bar{\tau}-1} <$$

$$\int_{Z^{t,t+\bar{\tau}-1}(m,m',s^{t,t+\bar{\tau}-1})} [2(1-\beta)][\mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1}) - \mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})]\, \mathrm{d}\, \gamma^{t,t+\bar{\tau}-1}$$

which implies

$$\int_{Z^{t,t+\bar{\tau}-1}(m,m',s^{t,t+\bar{\tau}-1})} [\mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1}) - \mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})]\, \mathrm{d}\, \gamma^{t,t+\bar{\tau}-1} > \frac{\eta}{4(1-\delta^T)(1-\beta)}$$

Consider now the payoffs in state $\omega = 1$ We have

$$\int_{Z^{t,t+\bar{\tau}-1}(m,m',s^{t,t+\bar{\tau}-1})} |[1-\mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})]^2 - [1-\mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})]^2| \;\; d\, \gamma^{t,t+\bar{\tau}-1} =$$

$$\int_{Z^{t,t+\bar{\tau}-1}(m,m',s^{t,t+\bar{\tau}-1})} [1-\mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})]^2 - [1-\mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})]^2 \;\; \mathrm{d}\, \gamma^{t,t+\bar{\tau}-1} =$$

$$\int_{Z^{t,t+\bar{\tau}-1}(m,m',s^{t,t+\bar{\tau}-1})} [\mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1}) - \mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})]$$

$$[2-\mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1}) - \mathrm{a}_{t+\bar{\tau}}(m', s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})]\, \mathrm{d}\;\gamma^{t,t+\bar{\tau}-1} \geq$$

$$2\beta \int_{Z^{t,t+\bar{\tau}-1}(m,m',s^{t,t+\bar{\tau}-1})} [\mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1}) - \mathrm{a}_{t+\bar{\tau}}(m, s^{t,t+\bar{\tau}-1}, \gamma^{t,t+\bar{\tau}-1})] \;\; \mathrm{d}\, \gamma^{t,t+\bar{\tau}-1} > \frac{2\beta\eta}{4(1-\delta^T)(1-\beta)}$$

The strict inequality between the first and the last expression above, together with Step 1, implies that the difference $U_t(m, 1) - U_t(m', 1)$ is bounded below by

$$\frac{\beta\eta(1-\delta)\delta^T(1-p)^T}{2(1-\delta^T)(1-\beta)}$$

which contradicts Step 2. Hence the proof of Lemma A.2 is now complete.